

**DEVELOPMENT OF A WAVEFORM SAMPLING ASIC WITH
FEMTOSECOND TIMING FOR A LOW OCCUPANCY VERTEX
DETECTOR**

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII AT MĀNOA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ELECTRICAL ENGINEERING

MARCH 2018

By

Peter Orel

Dissertation Committee:

Gary S. Varner, Chairperson

Victor M. Lubecke

David G. Garmire

Zhengqing Yun

Sven Vahsen

My work is dedicated to my *wife*

Irena Orel

for being patient and giving me the needed support to see this through.

I could not have done this without you.

Special thanks to my family and friends who helped me become who I am today.

My parents:

Janja and Dušan Orel,

My brother:

Borut Orel.

Thanks to my co-workers and friends at Instrumentation Technologies who made an engineer out of me.

ACKNOWLEDGMENTS

I would like to acknowledge the University of Hawai‘i at Mānoa (UH Mānoa), specifically the Instrumentation Development Laboratory (IDLab), for supporting this work. I would also like to thank my research advisor, Prof. Gary S. Varner, for giving me the opportunity to work and grow as a scientist. I would like to thank Dr. Andrej Seljak for his continued guidance and support during the various publications processes. Many thanks to Harley Cumming who helped me considerably with the Cadence software.

I have been very fortunate to be surrounded by knowledgeable and helpful individuals working in the IDLab. These people include Bronson Edralin, James Bynes, Khanh Le, Matthew Andrew, Lauri Vihtori Virta and everyone else. I would also like to take the time to thank the members of my dissertation committee: Prof. Victor M. Lubecke, Prof. David G. Garmire, Prof. Zhengqing Yun, and Prof. Sven Vahsen for their time, and guidance.

ABSTRACT

Vertex detectors provide space-time coordinates for the traversing charged particle decay products closest to the interaction point of a high-energy particle collider. Resolving these increasingly intense particle fluences at higher luminosities (larger number of collisions per second) is an ever-growing challenge. Furthermore, such fluences result in a non-negligible occupancy of the vertex detectors using existing low material budget techniques. Consequently, new approaches are being studied that meet the vertexing requirements while lowering the occupancy and the data rate. In this work we introduce the architecture and specifications for a novel vertex detector design based on femtosecond precision timing. The feasibility study results indicate that the new detector ladder design could achieve an occupancy ten times lower than its predecessor in the Belle II spectrometer, while maintaining a comparable spatial resolution. Furthermore, this leads to a considerable reduction in the detector data rate, thus lowering the cost of the subsequent processing electronics. One of the crucial parts of the detector is its readout ASIC (RFpix), whose development steps are discussed in detail. The RFpix is a twelve-bit resolution waveform digitizer with a sampling speed of 20 GS/s and an analog bandwidth of 3 GHz. Post-layout simulation results of the RFpix prototype analog front-end are shown and thoroughly analyzed. The simulated performance is shown to match the RFpix requirements, thus reaching an exquisite timing resolution of 160 fs.

CONTENTS

Acknowledgments	iv
Abstract	v
List of Tables	x
List of Figures	xi
Preface	xxv
1 Introduction	1
1.1 The Intensity Frontier	1
1.2 SuperKEKB and The Belle II Experiment	3
1.2.1 SuperKEKB	4
1.2.2 The Belle II Experiment	5
1.3 Vertex Detectors	7
1.3.1 The Belle II PXD	9
1.3.2 Summary of Vertex Detectors	11
2 Timing Vertex Detector	12
2.1 Timing Vertex Detector (TVD) Sensor Ladder Architecture	12
2.2 Principle of Operation	13
2.3 Feasibility Study, Requirements, Results, and Technical Challenges	21
2.3.1 Design Requirements	21
2.3.2 Transmission Line	24
2.3.3 The Pixel	28
2.3.4 The Readout application specific integrated circuit (ASIC)	30

2.3.5	Initial TVD Subpart Design Specifications and Summary	33
3	PSEC4 Analysis Results	36
3.1	Basic Design Scheme and Layout	37
3.2	DC Analysis and Characterization of Basic Components	38
3.2.1	Switch Resistance	39
3.2.2	Sampling Cell Capacitances	39
3.3	Frequency Response of the PSEC4 Sampling Cell	41
3.3.1	Tracking Bandwidth	41
3.3.2	Group Delay	41
3.4	Large Signal Analysis	41
3.5	Noise and ENOB	43
3.5.1	Noise	43
3.5.2	ENOB	44
3.6	Transient Response	46
3.7	Summary of PSEC4 Sampling Cell Analysis	49
4	Femtosecond Resolution Timing in Multi-GS/s Waveform Digitizing ASICs . .	50
4.1	Waveform Sampling Technique	50
4.2	Synthetic Waveform Generator	50
4.3	Case Studies and Analysis	53
4.4	Simulation Results	54
4.4.1	Synchronous vs Asynchronous Mode and Signal Shape Impact	54
4.4.2	Amplitude Noise vs Quantization Effects	57
4.4.3	Jitter	60

4.4.4	Scaling of $\sigma_{t_{INT}}$ as a Function of Sampling Frequency, Bandwidth, and Noise	60
4.4.5	Timing Extraction Algorithm Comparison	63
4.4.6	Summary of Simulation Results	66
4.5	Test Bench Measurements	67
4.5.1	Test Bench Setup	67
4.5.2	Measurement Results	69
4.6	Alternative Waveform Shapes	78
4.7	Summary	79
5	RFpix Prototype Design	81
5.1	RFpix1 Specifications and Architecture	81
5.2	Switched Capacitor Array	85
5.2.1	Sampling Cell, First Stage (Sampling Switches and Sampling Capacitor) . . .	85
5.2.2	Sampling Cell, Second Stage (Sampling Cell Buffer)	88
5.2.3	Differential Clock Driver	95
5.2.4	Summary of the Switched Capacitor Array	96
5.3	Analog Storage Array	98
5.3.1	Analog Storage Cell	99
5.3.2	RC Wire Model	103
5.3.3	Analog Buffer Depth Estimation	104
5.3.4	Summary of Analog Storage Array	106
5.4	Two-Level Delay-Locked Loop	107
5.4.1	Jitter and its Propagation in Inverters and D Flip-Flops	107
5.4.2	Two-Level DLL	110

5.5	General Purpose Input/Output Rail-to-Rail Operational Amplifier	120
5.6	Summary of the RFpix1 Analog Front-End	126
6	Summary	129
	Bibliography	130

LIST OF TABLES

2.1	Summary and comparison of the Belle II Pixel Vertex Detector (PXD) and the TVD	17
2.2	Parametric values used in calculating the data rates for two different timing extraction algorithms with different numbers of waveform samples.	21
2.3	A summary of TVD sensor baseline design parameters.	24
2.4	Baseline TVD design specifications.	34
3.1	PSEC4 characteristics	36
3.2	Extrapolated PSEC4 sampling cell capacitances.	41
3.3	PSEC4 acquisition and settling times at three different input V_{DC} voltages.	48
4.1	$\sigma_{t_{INT}}$ in the synchronous and asynchronous modes of operation for all three signal types.	54
4.2	Test bench equipment.	68
4.3	Test bench equipment performance parameters.	69
4.4	Summary and comparison of oscilloscope measurements and synthetic waveform generator (SWG) simulations results.	70
5.1	Baseline RFpix1 design specifications.	82
5.2	Simulated RFpix1 performance in comparison with the baseline specifications. . . .	126

LIST OF FIGURES

1.1	Research in high energy physics is pushing forward in three frontiers: Energy, Intensity, and Cosmic [1].	1
1.2	A concise summary of the constituents of the Standard Model of particle physics [2].	3
1.3	SuperKEKB particle accelerator with the Belle II spectrometer shown at the upper right [3].	5
1.4	The Belle II spectrometer is composed of seven sub-detectors and it measures 7.5 <i>m</i> in length and 7 <i>m</i> in height [3].	7
1.5	(a) Illustration of particle decay mechanism along with the corresponding vertices and tracks in the IP (side view) [4]. (b) Illustration of particle decay mechanism along with the corresponding vertices and tracks in the IP (beam side view) [4]. . . .	8
1.6	(a) Principle of operation of a silicon strip detector [4]. (b) Illustration of a vertex detector sensor ladder [4].	9
1.7	(a) Schematic view of the geometrical arrangement of the sensors for the PXD. The light grey surfaces are the sensitive DEPLETED Field Effect Transistor (DEPFET) pixels thinned to 50 μm and covering the entire acceptance of the tracker system. The full length of the outer modules is 174 <i>mm</i> [5]. (b) Operating principle and structure of a DEPFET [6].	10
1.8	The PXD data volume in comparison with the rest of the spectrometer.	11
2.1	3D sketch of the proposed TVD sensor ladder architecture.	13
2.2	Voltage pulse generation and injection into the transmission line (top). Pulse propagation in the transmission line (center). Measurement of the pulse arrival time (bottom).	14
2.3	Propagation delay and propagation speed as functions of the substrate ϵ_r for pixels close to the readout ASIC with a sensor ladder length of 90 <i>mm</i>	15
2.4	Spatial resolution as a function of propagation speed and timing resolution.	16
2.5	Occupancies of the Belle II PXD and the proposed TVD as functions of hit rate density.	18

2.6	Occupancy at a $\rho_{hit\ rate} = 4.62 \cdot 10^5\ Hits/cm^2s$, 1 MHz dark count rate per channel, and spatial resolution at timing resolution of 100 fs as functions of ϵ_r	18
2.7	Probability of multiple event hits per channel per time window at the worst-case hit rate.	19
2.8	(a) Acceptance area over pixel total area ratio as a function of the pixel active area along the sensor ladder. (b) Overall increase in occupancy as a function of pixel active area thickness.	20
2.9	Timing resolution parameter sweeps in terms of f_{SAMP} and Δu (top), BW and Δu (center), and f_{SAMP} and BW (bottom) with clear indications of the desired operating region (lightly shaded area) and the nominal operating points (OP). . . .	23
2.10	Direct relations between the design requirements and the corresponding subparts. . .	24
2.11	Voltage pulse approximation on the transmission line.	25
2.12	Cross section of the transmission line geometry with the vertical dimensions given in brackets.	26
2.13	(a) Transmission line loss for the worst-case signal travel path of 19.6 cm . (b) Timing resolution as a function of signal power.	27
2.14	Measurement results of phase drift in cables over time with clear temperature dependence.	27
2.15	Single channel block diagram of the far-end termination ASIC with directional coupling and amplification.	28
2.16	(a) Simplified block diagram of a single pixel cell. (b) Simplified equivalent circuit of the pixel driver, through silicon via (TSV), transmission line segment, and matching circuit.	29
2.17	(a) Non-matched and matched single pixel insertion loss as a function of frequency. (b) Simulation results of a transmission line loaded with 1600 pixels.	30
2.18	RFpix readout ASIC functional block diagram.	32
3.1	Block diagram of a single PSEC4 SCA sampling cell.	37
3.2	(a) PSEC4 input channel layout with input pad, transmission line, and sampling cell. (b) Layout of a single PSEC4 sampling cell showing the switch driver, input CMOS switch, sampling capacitor, and the transmission line tap.	38

3.3	Single PSEC4 sampling cell with equivalent circuit.	38
3.4	Track mode (ON-state) and Hold mode (OFF-state) resistances as functions of input V_{DC}	39
3.5	Extrapolated capacitance as a function of frequency.	40
3.6	Bandwidth and track mode switch resistance as functions of Input V_{DC}	42
3.7	Sampling cell group delay as a function of input V_{DC} and frequency.	43
3.8	(a) PSEC4 sampling cell signal attenuation as a function of input V_{DC} and signal amplitude at a signal frequency of 1 GHz . (b) PSEC4 sampling cell signal attenuation as a function of frequency and signal amplitude at an input V_{DC} of 650 mV . (c) Total harmonic distortion as a function of input V_{DC} and signal amplitude. . . .	44
3.9	Distortion and signal attenuation as functions of signal amplitude at input $V_{DC} = 600 V$	45
3.10	Output referred noise as a function of input V_{DC} and frequency.	46
3.11	(a) ENOB of the PSEC4 sampling cell as a function of input V_{DC} and signal amplitude. (b) ENOB of the PSEC4 sampling cell as a function of frequency and signal amplitude.	47
3.12	PSEC4 sampling cell transient response with 1 ns time window at three different input V_{DC} voltages: (a) $V_{DC} = 300 mV$, (b) $V_{DC} = 600 mV$, and (c) $V_{DC} = 900 mV$	48
4.1	Functional block diagram of the SWG code with the time base generator and the waveform synthesizer substructures. Input variables are marked in black on the left side, while the output vectors are marked in yellow on the right side.	52
4.2	Example of a noisy Gaussian pulse waveform generated by the SWG with an overlay visualizing the linear extraction algorithm.	53
4.3	Synchronous (left) and asynchronous (right) modes of operation ($N_{acq.w.} = 30$). . . .	55
4.4	Persistence plot of 500 fitted waveforms over the transition point in the synchronous mode.	56
4.5	Interpolated transition time resolution as a function of the mean value of the sampling interpolation fraction.	57

4.6	(a) Simulated and estimated interpolated transition time resolution as a function of N_{BITS} . (b) Signal spectrum for sampled signals for three different vertical resolutions in terms of N_{BITS}	58
4.7	(a) Histograms of simulated interpolated transition times for the 8-bit (left) and 12-bit (right) resolution cases respectfully. (b) Interpolated transition time resolution as a function of noise and N_{BITS}	59
4.8	Interpolated transition time resolution along with the corresponding μ_{SIF} as functions of signal amplitude. Each signal voltage point has been simulated with a population of 1000.	59
4.9	Interpolated transition time resolution as a function of added jitter at different numbers of bits.	60
4.10	Simulated interpolated transition time resolution (left) and the corresponding μ_{SIF} (right) as functions of noise and sampling speed.	61
4.11	Interpolated transition time resolution as a function of sampling speed and bandwidth with spots of low $\sigma_{t_{INT}}$	62
4.12	Asymptotic interpolated transition time resolution as a function of sampling speed. .	63
4.13	Timing resolution as a function of number of samples by using different extraction algorithms for both, synchronous and asynchronous cases.	64
4.14	Timing resolution as a function of the Sample Interpolation Factor (SIF) by using different extraction algorithms (synchronous mode) on a sine waveform.	65
4.15	(a) Timing resolution as a function of the SIF by using different extraction algorithms (synchronous mode) on a square pulse waveform. (b) Timing resolution as a function of the SIF by using different extraction algorithms (synchronous mode) on a Gaussian pulse waveform.	66
4.16	Signal period extraction in time domain with the corresponding SIF.	68
4.17	Test bench setup block diagram (asynchronous mode).	69
4.18	(a) Comparison between the OSC1 test data and the corresponding simulation in terms of the extracted signal period resolution as a function of μ_{SIF} . (b) Distributions of the extracted signal period resolution (left), μ_{SIF} (center), and σ_{SIF} (right) for the OSC1 test and the corresponding simulation with $f_{SIG} = 6\text{ GHz}$	71
4.19	Comparison between the OSC1 test data and the corresponding simulation in terms of the extracted signal period resolution as a function of SIF.	72

4.20	Sampled signal waveform in time domain at signal frequencies of 5.7143 GHz ($f_{SAMP}/f_{SIG.} = 7$) and 6 GHz ($f_{SAMP}/f_{SIG.} = 6.667$).	72
4.21	Comparison between the OSC3 test data and the corresponding simulation in terms of the extracted signal period resolution as a function of μ_{SIF}	73
4.22	(a) Comparison between analog to digital converter (ADC) test data and the corresponding simulation in terms of the extracted signal period resolution as a function of μ_{SIF} . (b) Extracted signal period resolution from the ADC data as a function of f_{SAMP} and $f_{SIG.}$	74
4.23	(a) Extracted signal period resolution $\sigma_{T_{PERIOD}}$ from the ADC data as a function of μ_{SIF} for available ADC sampling speeds at a signal frequency of 500 MHz. (b) Extracted signal period resolution $\sigma_{T_{PERIOD}}$ from the ADC data as a function of μ_{SIF} for signal frequencies ranging from 400 MHz to 1 GHz at the ADC sampling speed of 4 GS/s.	75
4.24	(a) Extracted signal period resolution from the ADC data as a function of RSD_{SIF} for the available ADC sampling speeds at the signal frequency of 500 MHz. (b) Measured (colored points) and simulated (black points) signal period resolution as a function of $relativestandarddeviation(RSD)_{SIF}$ for signal frequencies ranging from 400 MHz to 1 GHz at an ADC sampling speed of 4 GS/s.	75
4.25	Measured and simulated signal period resolution as a function of the $f_{SAMP}/f_{SIG.}$ ratio for signal frequencies ranging from 400 MHz to 1 GHz at an ADC sampling speed of 4 GS/s.	76
4.26	Test bench setup block diagram (synchronous mode).	77
4.27	Measured and simulated $\sigma_{T_{PERIOD}}$ as a function of μ_{SIF} for the synchronous mode of operation.	77
4.28	(a) Time domain plot of a Ricker wavelet with a rise time of 120 ps and an approximate duration of 1.1 ns. (b) Single sideband spectrum of the Ricker wavelet with the maxima and 3 dB points displayed.	78
5.1	Simplified functional block schematic of the RFpix1.	82
5.2	(a) Simplified RFpix1 sampling cell schematic showing the differential input structure with the input switch, sampling capacitor, the differential clock buffer, and the sampling cell buffer, which is an indirect current-feedback instrumentation amplifier. (b) RFpix1 sampling cell track-mode switch resistance and tracking bandwidth as functions of input V_{DC}	85

5.3	(a) Comparison between the RFpix1 and the PSEC4 sampling cell track-mode switch resistances as functions of input V_{DC} . (b) Comparison between the RFpix1 and the PSEC4 sampling cell tracking bandwidths as functions of input V_{DC}	86
5.4	(a) RFpix1 sampling cell THD curves as functions of signal amplitude at different input signal frequencies. (b) Input and sampling cell capacitances in track and hold mode as functions of input DC voltage.	87
5.5	(a) Track mode single sideband (SSB) noise spectrum of the sampling cell. (b) Sampling cell integrated noise floor as a function of sampling capacitance.	87
5.6	Instrumentation amplifier realization with three op-amps.	88
5.7	Schematic showing an indirect current-feedback instrumentation amplifier. The amplifier is implemented using bipolar transistors, but the same principle applies to a CMOS variant [7].	89
5.8	Transistor level schematic of the sampling cell buffer.	91
5.9	(a) Sampling cell buffer DC response at three different temperatures. (b) Sampling cell buffer transient response.	92
5.10	(a) Sampling cell buffer small signal bandwidth as a function of input DC voltage. (b) Sampling cell buffer gain and phase at an input DC voltage of 600 mV	93
5.11	(a) Pole diagram of the sampling cell buffer for two input DC offsets. (b) Sampling cell buffer total harmonic distortion as a function of input signal amplitude.	93
5.12	(a) Sampling cell buffer noise spectrum. (b) Sampling cell buffer common mode rejection ratio.	94
5.13	(a) Sampling cell buffer quiescent current draw as a function of input voltage. (b) Sampling cell buffer transient current draw for three different output voltage amplitude swings.	94
5.14	Sampling cell buffer layout.	95
5.15	Differential clock driver propagation delay as a function of temperature.	95
5.16	(a) Simplified schematic of the switched capacitor cell. (b) Timing diagram of the switched capacitor cell strobe signals.	96
5.17	Switched capacitor cell layout.	97

5.18	(a) Sampled single-ended sampling cell signals with the corresponding reconstructed signals. (b) Original and reconstructed signals as functions of single-ended input signal amplitude.	97
5.19	Storage array functional schematic portraying the row-column format of the storage cells.	98
5.20	Storage cell functional schematic.	99
5.21	Storage cell ON-state, OFF-state, and storage capacitances as functions of input voltage at various temperatures.	100
5.22	(a) Switch ON-state resistance as a function of input voltage at different temperatures. (b) Storage cell tracking as a function of input voltage at different temperatures.	101
5.23	(a) Total harmonic distortion as a function of signal amplitude at different temperatures and signal frequency of 312.5 MHz. (b) Total harmonic distortion as a function of frequency at the signal amplitude of 1 V_{PP}	102
5.24	(a) Comparator propagation delay as a function of storage cell voltage at three different temperatures. (b) Cumulative storage cell error (from charge injection, propagation delay, and threshold variation) as a function of the storage cell voltage.	103
5.25	Storage cell layout.	103
5.26	(a) Capacitance and resistance for the wire connecting the sampling and storage arrays as functions of length for four different wire widths (b) Bandwidth of the wire connecting the sampling and storage arrays as a function of length for four different wire widths.	104
5.27	Buffer depth simulation flow chart.	105
5.28	RFpix1 analog storage histogrammed buffer depth per time window.	106
5.29	Functional block schematic of a DLL.	108
5.30	Added jitter of a single inverter as a function of input rise time and current consumption for a clock frequency of 156 MHz (a) and 312.5 MHz (b).	109
5.31	Inverter added jitter as a function of temperature.	109
5.32	Added jitter (a) and rise time (b) along the delay line as functions of DLL cell number.	110
5.33	(a) Simplified block schematic of the modified fully differential D flip-flop. (b) Added jitter as a function of temperature at the outputs of three cascaded DDFFCs.	111

5.34	(a) Added jitter as a function of DLL cell number and temperature (b) Jitter margin to 100 fs as a function of added jitter.	112
5.35	Comparison between DLL architectures in terms of power consumption and added jitter.	112
5.36	Simplified block schematic of a two-level cascaded DLL.	113
5.37	(a) L1DL cell to cell delay as a function of temperature with a $\pm\sigma$ mismatch estimate at 30 $^{\circ}C$. (b) Layout of the modified RFpix1 DDFFC cell.	114
5.38	Timing diagram of the designed two-level DLL. The time scale used in the figure is in ns	115
5.39	Pseudo differential starved inverter layout.	116
5.40	Simulated L2 cell delay (a), L2 delay line added jitter (b) as functions of temperature, and the cumulative delay deviation of the L2 delay line due to mismatch (c).	117
5.41	Simulated delay as a function of control voltages for both, the rising (a) and the falling edges (b).	118
5.42	L2DL feedback control loop block schematic.	118
5.43	Smoothed L2DL response as a function of control voltage.	119
5.44	Phase error response of the L2DL array in time.	119
5.45	Snapshot of the two-level DLL with important parts highlighted.	120
5.46	Schematic symbols of four variations of the general purpose input/output rail-to-rail operational amplifier implementation. (a) self-biased, (b) unity gain self-biased, (c) tunable biases (current source and g_m compensation), (d) tunable frequency response.	121
5.47	Schematic of the general purpose input/output rail-to-rail operational amplifier implementation.	122
5.48	Layout of the general purpose input/output rail-to-rail operational amplifier implementation.	123
5.49	(a) Input/output characteristic as a function of the input DC voltage in a unity gain buffer configuration. (b) Transconductance as a function of input DC voltage.	124
5.50	(a) Open-loop gain and phase response of the op-amp. (b) Open-loop phase margin as a function of load capacitance and temperature.	124

5.51 (a) Closed-loop phase margin of the op-amp in a unity gain configuration as a function of load capacitance and temperature. (b) Closed-loop large signal bandwidth as a function of load capacitance and temperature for the unity gain op-amp configuration. 125

5.52 (a) Amplifier input referred noise for six different temperatures. (b) Amplifier integrated noise floor as a function of temperature and load capacitance. 125

5.53 RFpix1 analog front-end layout. 127

5.54 Total estimated sampling error of the RFpix1 analog front-end. 128

LIST OF TERMS

A_{TH} . Threshold voltage.

A_i, A_{i+1} Amplitudes of samples adjacent to the S_{INT} .

BW Bandwidth.

T_{SAMP} . Sampling period.

f_{SAMP} . Sampling speed.

f_{SIG} . Frequency of the signal.

J_{ADDED} Added jitter.

J_{INJ} . Injected jitter.

N_{BITS} Number of bits.

$N_{acq.w.}$ Number of acquisition windows.

$N_{periods}$ Number of periods per acquisition window.

RSD_{SIF} . Relative standard deviation of the sample interpolation fraction.

S_{INT} . Interpolated threshold transition point.

S_i, S_{i+1} Samples adjacent to the S_{INT} .

t_{INT} . Interpolated transition time through the chosen threshold voltage.

T_{PERIOD} Extracted signal period.

t_r Rise time.

t_i, t_{i+1} Time stamps of the samples adjacent to the S_{INT} .

U Signal amplitude.

Δu_{AMP} Amplitude noise.

Δu Noise.

Δu_Q Quantization noise.

μ_{SIF} Mean value of the SIF FOM within an acquisition window.

σ_{SIF} Standard deviation of the SIF FOM within an acquisition window.

$\sigma_{T_{PERIOD}}$ Resoltuion of the extracted signal period.

$\sigma_{t_{INT.}}$ Interpolated transition time resolution.

LIST OF ACRONYMS

ADC analog to digital converter.

AM-to-PM amplitude modulation to phase modulation.

APD avalanche photodiodes.

ASIC application specific integrated circuit.

CDGN clock distribution and generation network.

CV coefficient of variation.

CW continuous wave.

DAC digital to analog converter.

DAQ data acquisition.

DCR dark count rate.

DEPFET DEpleted Field Effect Transistor.

DLL delay-lock loop.

FOM figure of merit.

FPGA field programmable gate array.

FWHM full width at half maximum.

GaAs Gallium Arsenide.

GUI graphical user interface.

HIresSi High-resistivity Silicon.

IC integrated circuit.

IP interaction point.

JFET junction field effect transistor.

LAPPD Large Area Picosecond PhotoDetectors.

LGAD low gain avalanche photodiodes.

LSB least significant bit.

LSQ least-mean-square.

OP operating point.

OR operating region.

OSC oscilloscope.

PCB printed circuit board.

PXD Pixel Vertex Detector.

R&D research and development.

RF radio frequency.

RMS root mean square.

RSD relative standard deviation.

SCA switched capacitor array.

SIF Sample Interpolation Factor.

SiPM silicon photomultiplier.

SNR signal-to-noise ratio.

SPI serial peripheral interface.

SR slew rate.

SWG synthetic waveform generator.

THD total harmonic distortion.

TOF time-of-flight.

TSV through silicon via.

TVD Timing Vertex Detector.

VDS vertex detector sensor.

XO crystal oscillator.

PREFACE

We begin by providing a brief introduction to the physics that drives the motivation to build and run high energy physics experiments. More specifically, we focus on the Belle II experiment and its vertex detector. Subsequently, the proposal for the novel vertex detector upgrade, denoted as the Timing Vertex Detector (TVD), is presented. The proposal covers the architecture of the TVD and its principle of operation. In addition, it documents the results of a detailed feasibility study, where the requirements and the technical challenges are comprehensively identified. A crucial part of the detector is its readout ASIC, named the RFPix, which needs to sample at a very high speed of 20 GS/s with a timing resolution of 100 fs. The baseline for the development of the RFPix is a waveform digitizer called the PSEC4, whose architecture is based on switched capacitor arrays. A thorough analysis of the PSEC4 sampling cell is presented, revealing the crucial sub-circuits that need to be improved to reach the RFPix specifications. Furthermore, simulation results that clearly identify and quantify the sources of error and the underlying coupling mechanisms when femtosecond timing is desired are shown. In addition, a synthetic waveform generation software tool developed solely for this purpose is presented and validated through measurement results, which provided vital knowledge, insights, and confidence for the development of the RFPix prototype or any other fast, wideband radio frequency (RF) system, that aims to achieve such exquisite timing performance. The RFPix prototype architecture and specifications are presented next, followed by an in-depth discourse on the development steps and post-layout simulation results of the RFPix prototype analog front-end. In addition, a direct comparison between the desired and the simulated performance of the ASIC is given. Finally, this work concludes with the summary of the key points of this research.

CHAPTER 1

INTRODUCTION

In this chapter we briefly introduce the Intensity Frontier, which is one of the three main areas of research in high energy physics. Particle physics experiments at the Intensity Frontier explore fundamental particles and forces of nature using intense particle beams and highly sensitive detectors. The Belle II spectrometer is a high energy physics experiment searching for rare phenomena at the Intensity Frontier. The general layouts of the collider and spectrometer are shown and the main performance parameters are briefly discussed in the following sections. The focus of the chapter is on the vertex detectors and their role in spectrometers. Furthermore, the Belle II pixel vertex detector (PXD) is introduced and discussed. The PXD is used as the baseline for comparison with the proposed Timing Vertex Detector (TVD) throughout chapter two.

1.1 The Intensity Frontier

Particle physics aspires to comprehend the fundamental laws that govern the universe around us. The Standard Model of particle physics describes the basic structures of matter and forces to the extent we have been able to probe thus far. Despite its overwhelming success in describing and predicting fundamental interactions since its inception, it leaves some big questions unanswered. Some are within the Standard Model itself, such as why there are so many fundamental particles and why they have different masses; in other cases, the Standard Model simply fails to explain some phenomena, such as the observed matter-antimatter asymmetry in the universe, the existence of dark matter and dark energy, and the mechanism that reconciles gravity with quantum mechanics. These gaps lead us to conclude that the universe must contain new and unexplored elements of nature. Particle physics is directed towards discovering and understanding these new laws of physics.

These questions are best pursued with a variety of approaches rather than with a single experiment or a technique. Particle physics uses three basic approaches often characterized as exploration along the Cosmic, Energy, and Intensity Frontiers, as illustrated in Figure 1.1. Each employs different tools and techniques, but ultimately they address the same fundamental questions. This allows a multi-pronged approach where attacking basic questions from different angles furthers knowledge

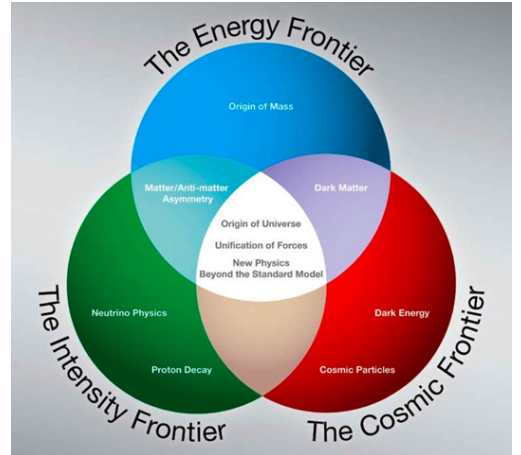


Figure 1.1: Research in high energy physics is pushing forward in three frontiers: Energy, Intensity, and Cosmic [1].

and provides deeper answers, so that the whole is more than a sum of parts. A coherent picture or underlying theoretical model can more easily emerge to be proven either correct or not.

The Intensity Frontier explores fundamental physics with intense sources and ultra-sensitive, sometimes massive, detectors. It encompasses searches for extremely rare processes and for tiny deviations from the Standard Model expectations. Intensity Frontier experiments use precision measurements to probe quantum effects. They typically investigate new laws of physics at energies higher than the kinematic reach of the high energy particle accelerators [1]. The science addresses basic questions, such as:

- Are there new sources of charge parity (CP) violation? A CP violation is a violation of the CP-symmetry: the combination of C-symmetry (charge conjugation symmetry) and P-symmetry (parity symmetry). CP-symmetry states that the laws of physics should be the same if a particle is interchanged with its antiparticle (C-symmetry) while its spatial coordinates are inverted ("mirror" or P-symmetry) [4].
- Is there a CP violation in the leptonic sector? A lepton is an elementary particle that does not undergo strong interactions. The best known lepton is the electron. There are six types of leptons, known as flavours, forming three generations. The first generation are the electronic leptons, comprising the electron (e^-) and electron neutrino (ν_e); the second are the muonic leptons, comprising the muon (μ^-) and muon neutrino (ν_μ); and the third are the tauonic leptons, comprising the tau (τ^-) and the tau neutrino (ν_τ). The leptonic sector refers to all six elementary particles [8].
- Are neutrinos their own antiparticles? As postulated by Paul Dirac, a particle and its antiparticle have the same mass as one another but opposite electric charge, and other differences in quantum numbers. However, in 1937 Ettore Majorana hypothesized the existence of particles that are their own antiparticles. Antiparticles are collectively called antimatter.
- Do the forces unify? There are four known forces of nature: the gravitational force, the electromagnetic force, the weak force, and the strong force. The gravitational force is currently best described by the theory of general relativity. The other three forces are well described by the Standard Model of particle physics. Particles of the Standard Model are shown in Figure 1.2. As an example of unification, it has been proven that the weak and the electromagnetic force unify into the electroweak interaction at sufficiently high enough energies.
- Is there a weakly coupled hidden sector that is related to the dark matter? The weak interaction is one of the four fundamental forces of nature. Dark matter is a hypothetical type of matter distinct from ordinary matter because it does not appear to interact through the electromagnetic or strong interactions. It has never been directly observed; however, its existence would explain a number of otherwise puzzling astronomical observations [4].

- Are there undiscovered symmetries? In physics, a symmetry of a physical system is a physical or mathematical feature of the system (observed or intrinsic) that is preserved or remains unchanged under some transformations. Arguably the most important example of a symmetry in physics is that the speed of light has the same value in all frames of reference [1].
- Why is there more matter than antimatter in the universe?

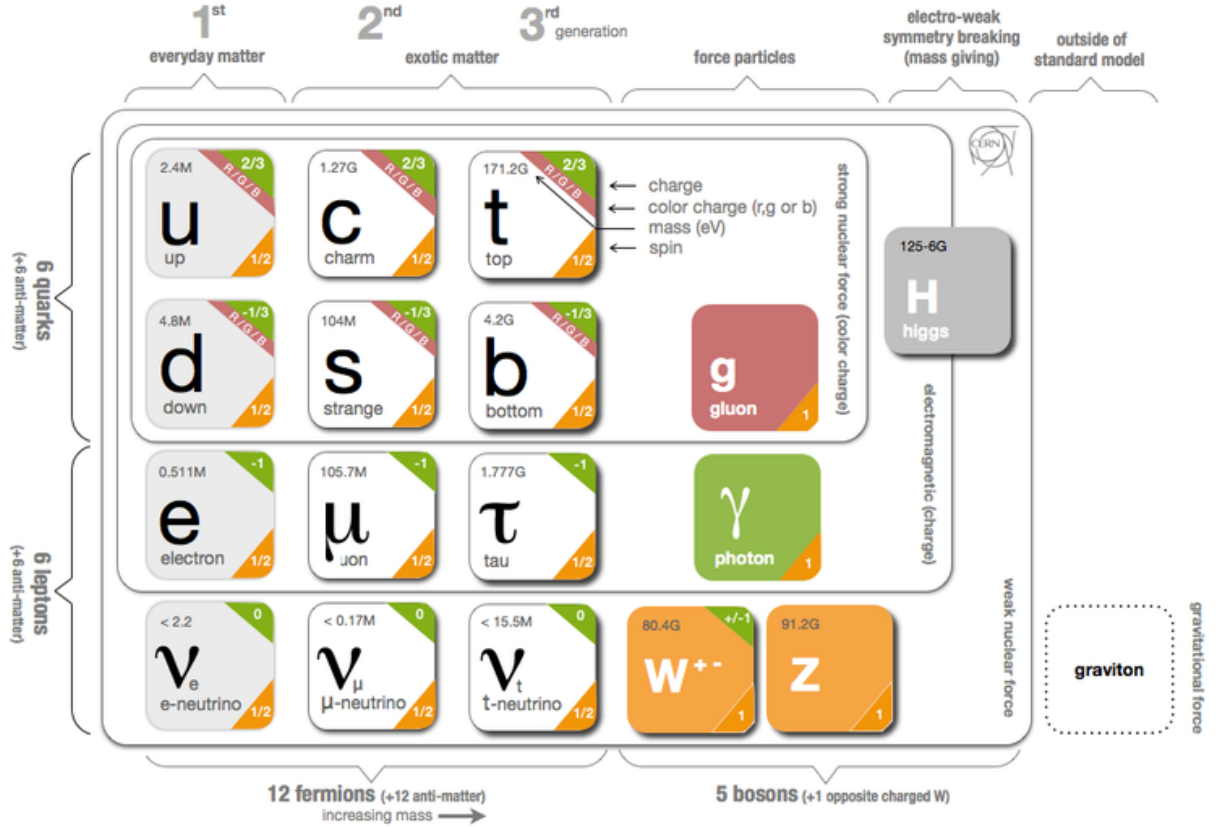


Figure 1.2: A concise summary of the constituents of the Standard Model of particle physics [2].

1.2 SuperKEKB and The Belle II Experiment

In order to serve answers to all of these questions and perhaps to formulate new ones, various experiments have been designed and constructed all around the world. One of the most famous and recognized experiments is the Large Hadron Collider (LHC) in Geneva, Switzerland. However, the LHC is considered an Energy Frontier experiment. On the Intensity Frontier, one of the largest and most important experiments is the Belle II experiment at the SuperKEKB particle collider.

1.2.1 SuperKEKB

SuperKEKB is a particle accelerator located at KEK (High Energy Accelerator Research Organisation) in Tsukuba, Ibaraki Prefecture, Japan. It is designed to collide electrons with positrons, the antimatter partners of electrons. In addition, SuperKEKB is a second-generation B-factory for the Belle II experiment. A B-factory is a particle collider experiment designed to produce and detect an extremely large number of B mesons so that their properties and behaviour can be measured with small statistical uncertainty. B mesons are mesons composed of a bottom antiquark and either an up (B^+), down (B^0), strange (B_S^0) or charm quark (B_c^+) [4]. The driving motivation is to precisely measure the $B - \bar{B}$ oscillation to perhaps explain the excess of matter over antimatter in the universe. The center-of-momentum energy is close to the mass of the $\Upsilon(4S)$ resonance. The $\Upsilon(4S)$ resonance is the fourth excited S-wave state of the Υ meson, which is a quarkonium state formed by a bottom quark and its antiparticle [4]. Quarks and their antimatter counterparts (antiquarks) are fundamental constituents of matter. The most famous is perhaps the combination of two up quarks and one down quark, which forms a proton. The $\Upsilon(4S)$ resonance predominantly decays into pairs of B mesons, hence the term B-factory. The accelerator is an upgrade to the KEKB accelerator, which achieved “first turns” (first circulation of electron and positron beams) in February 2016. First collisions are expected in April 2018 [9].

The SuperKEKB design reuses many components from the KEKB. Under normal operation, SuperKEKB will collide electrons at 7 GeV with positrons at 4 GeV (compared to KEKB at 8 GeV and 3.5 GeV, respectively). The eV unit stands for electron volt and it is a standard unit of energy used in physics. It is defined as the energy given to an electron by accelerating it through 1 V of electric potential difference [10]. The center-of-momentum energy of the collisions will therefore be slightly above the $\Upsilon(4S)$ mass ($10.57 \text{ GeV}/c^2$). The asymmetry in the beam energy provides a relativistic Lorentz boost to the B meson particles produced in the collision. A Lorentz boost is a Lorentz transformation without any rotations. Lorentz transformations are coordinate transformations between two reference frames moving with constant velocity relative to each other. These transformations are used when the relative velocity between two frames reaches relativistic speed, or in other words, the velocity is comparable to the speed of light. The core advantage is that the Lorentz transformations preserve the space-time interval between the two events. Space-time is a unified framework that fuses the three dimensions of space and the one dimension of time into a single four dimensional continuum [11]. The direction of the higher-energy beam determines the “forward” direction. The target luminosity of the accelerator is $8 \cdot 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$, which is 40 times larger than KEKB.

Luminosity is the proportionality factor between the interaction cross section and the number of interactions per second. The interaction cross section is the probability that two particles will collide and react in a certain way or, in other words, how many collisions will result in the desired event in relation to the total number of events [12].

As with KEKB, SuperKEKB consists of two storage rings: one for the high-energy electron beam (the High Energy Ring, HER) and one for the lower energy positron beam (the Low Energy Ring, LER). The accelerator has a circumference of 3016 m with four straight sections and experimental halls in the center of each, named “Tsukuba”, “Oho”, “Fuji”, and “Nikko”. The Belle II experiment is located at the single interaction point in Tsukuba Hall [13,14]. Figure 1.3 shows the SuperKEKB particle accelerator with the Belle II spectrometer at the upper right.

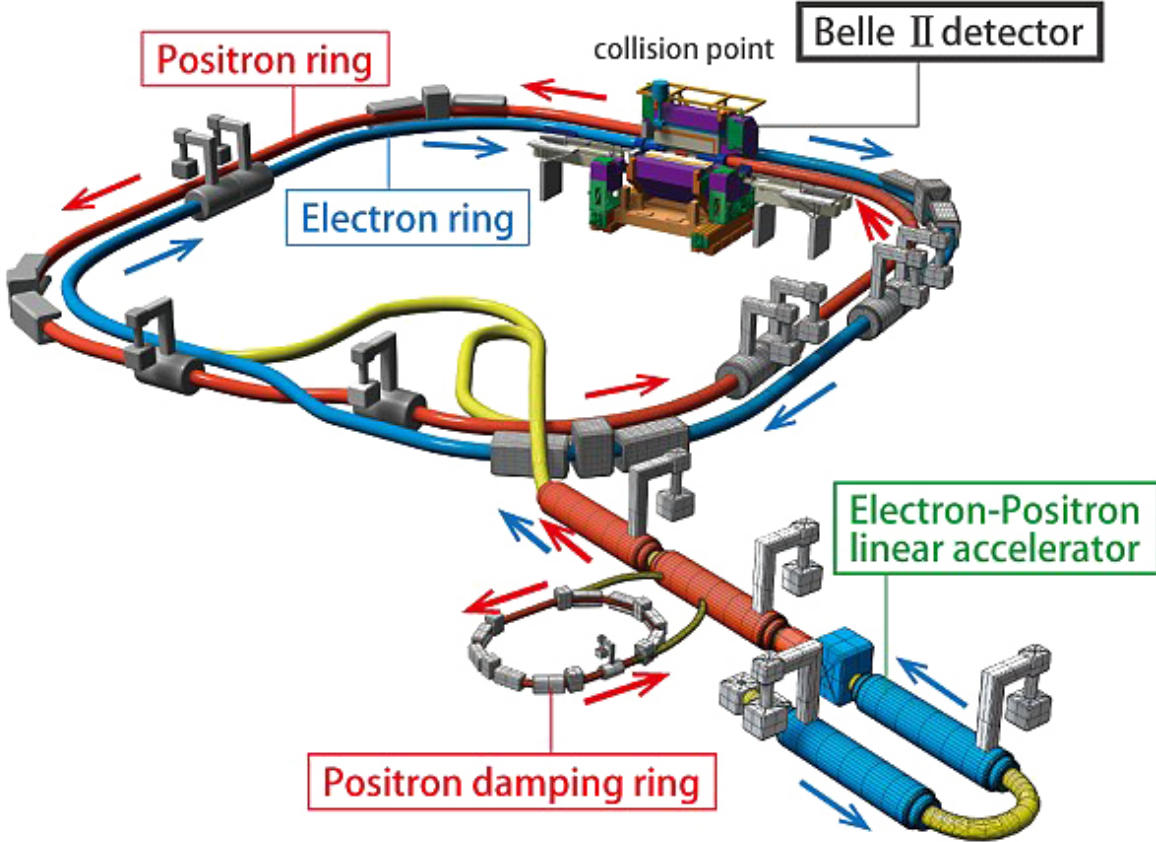


Figure 1.3: SuperKEKB particle accelerator with the Belle II spectrometer shown at the upper right [3].

1.2.2 The Belle II Experiment

The Belle II spectrometer is an upgrade to the Belle experiment, which took data from 1999 to 2010. Much of the original Belle spectrometer has been upgraded to cope with the higher instantaneous luminosity provided by the SuperKEKB accelerator. The Belle II spectrometer is composed of seven sub-detectors [3]. The following list enumerates the Belle II detectors in the approximate order in which they follow each other from the interaction point (IP) outwards:

1. The innermost layer is a two-layer pixel vertex detector (PXD) based on a depleted field effect transistors DEPFET. See the next section on vertex detectors for a more elaborate discussion on this topic.
2. The next layer outward from the PXD is the silicon vertex detector (SVD). The design of the Belle II SVD inherits the good characteristics of the Belle vertex detector: low mass, high precision, immunity to background hits, radiation tolerance, and long-term stability. It is designed with silicon strip sensors to avoid the massive channel count of pixels without compromising the vertex-detection capability of Belle II.
3. The next detection layer is the central drift chamber (CDC), which has three important roles. First, it reconstructs charged tracks and measures their momenta precisely. Second, it provides particle identification information using measurements of energy loss within its gas volume. Low momentum tracks, which do not reach the dedicated particle identification devices further outwards, can be identified using the CDC alone. Finally, it provides efficient and reliable trigger signals for charged particles.
4. In the barrel region of the spectrometer, we find the imaging Time-Of-Propagation (iTOP) counter. In this counter, the time of propagation of the Cherenkov photons internally reflected inside a quartz radiator is measured. The Cherenkov image is reconstructed from the three-dimensional information provided by two coordinates (x, y) and precise timing, which is determined by micro-channel plate (MCP) photomultiplier tubes (PMTs) at the backward surface of the quartz radiator assembly.
5. In the forward endcap, the proximity-focusing Aerogel Ring-Imaging Cherenkov detector (ARICH) has been designed to separate kaons from pions over most of their momentum spectrum and to provide discrimination between pions, muons, and electrons below $1\text{ GeV}/c$. A kaon or K mesons are bound states of a strange quark (or antiquark) and an up or down antiquark (or quark). Pions consist of up and down, quark and antiquark pairs. A muon is an elementary particle (lepton), like the electron, but with greater mass.
6. The original CsI(Tl) electromagnetic calorimeter (ECL) has been re-used, however, the read-out electronics have been upgraded. The main tasks of the calorimeter are: detection of photons with high efficiency, precise determination of the photon energy and angular coordinates, electron identification, generation of the proper signal for trigger, on-line and off-line luminosity, and K_L^0 detection together with the KLM measurement.
7. Finally, scintillators have been installed in the forward endcap and inner layers of Belle's K_L^0 and muon detector (KLM). The KLM consists of an alternating sandwich of 4.7 cm thick iron plates and active detector elements located outside the superconducting solenoid. The iron plates serve as the magnetic flux return for the solenoid [3].

Figure 1.4 shows a 3D cut-away diagram of the Belle II spectrometer with all seven sub-detectors shown in the diagram.

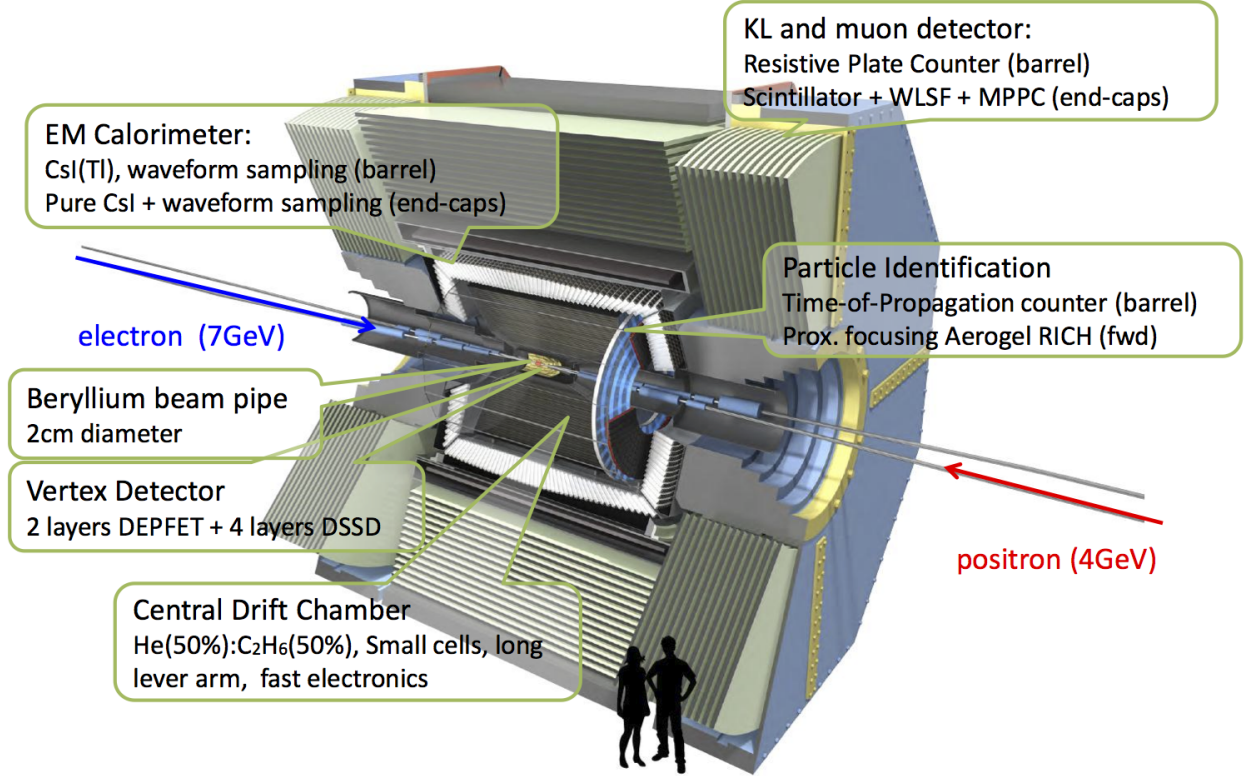


Figure 1.4: The Belle II spectrometer is composed of seven sub-detectors and it measures 7.5 *m* in length and 7 *m* in height [3].

1.3 Vertex Detectors

Vertex detectors are usually the closest detectors to the IP. They provide space-time coordinates for the traversing charged particle decay products closest to the IP. In the specific case of the Belle II spectrometer, the innermost layer of the vertex detector is situated a mere 14 *mm* away from the IP. Figure 1.5a shows an example of particles emerging from the IP and traversing the vertex detector. The particle origin positions are called vertices, hence the name vertex detector. The primary vertex is the collision point, where the new particles get generated. The secondary and tertiary vertices are the subsequent decay coordinates of the new particles originating from the IP, thus decaying to yet other new particles. This mechanism is illustrated in Figure 1.5b.

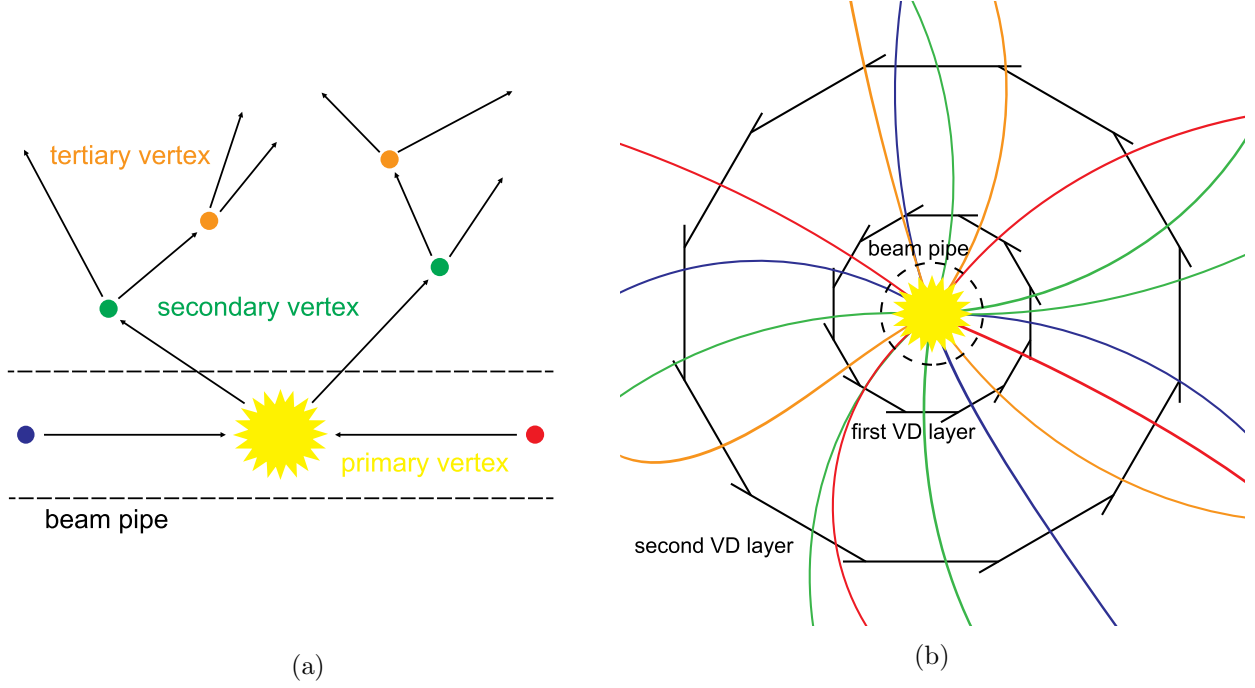


Figure 1.5: (a) Illustration of particle decay mechanism along with the corresponding vertices and tracks in the IP (side view) [4]. (b) Illustration of particle decay mechanism along with the corresponding vertices and tracks in the IP (beam side view) [4].

Due to their proximity to the IP, vertex detectors are subject to the highest hit rate densities and thus prone to very high occupancies. Occupancy of a detector is the ratio of the number of activated channels versus the total number of channels in a given time window. Particularly at high luminosities, most of the hits are from the background. The background is composed of the unwanted particles that are not a part of the primary generation process during the collision. Some of the possible background sources in Belle II are:

- Synchrotron radiation (SR) from the high energy ring (HER) upstream direction;
- Backscattering of SR from HER downstream;
- Scattering of the beam on residual gas;
- Touschek scattering (particle loss within a bunch due to a single particle-particle; collision)
- Radiative Bhabha scattering (electron-positron scattering process $(e^+e^- \rightarrow e^+e^-)$);
- Electron-positron pair production via the two photon process $(e^+e^- \rightarrow e^+e^-e^+e^-)$.

Furthermore, the background increases roughly with the inverse of the square of the distance from the IP [3].

Originally, most designs were based on the silicon micro-strip technology [15]. However, the relatively small number of channels and thus their susceptibility to higher occupancies makes them an impractical solution at high luminosities for the innermost vertexing layers. As a result, many current and future vertex detector designs are pixel based [16]. Figures 1.6a and 1.6b show illustrations of typical strip and pixel detectors respectively.

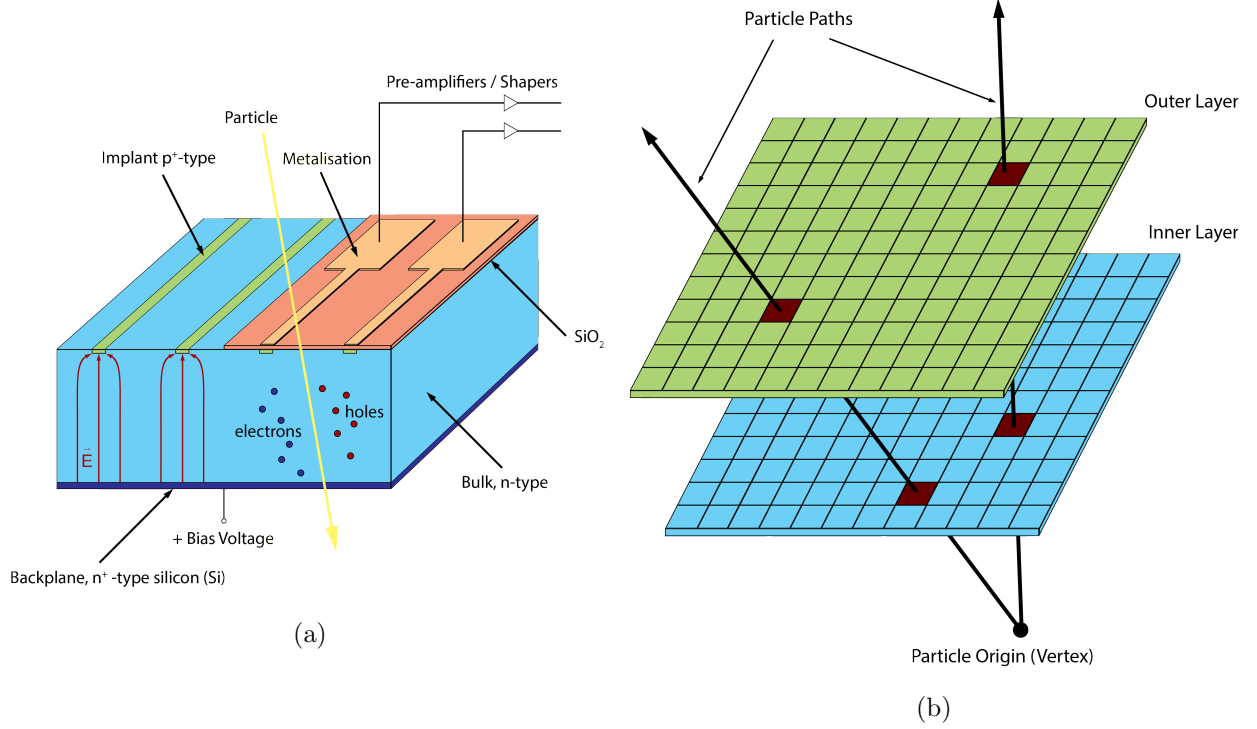


Figure 1.6: (a) Principle of operation of a silicon strip detector [4]. (b) Illustration of a vertex detector sensor ladder [4].

1.3.1 The Belle II PXD

Vertex detectors are composed of one or more vertex detector sensors (VDSs). The VDS architecture used at SuperKEKB for the two innermost layers relies on a pixel matrix with pixel dimensions of the order of $50\ \mu\text{m} \times 50\ \mu\text{m}$ and $50\ \mu\text{m} \times 75\ \mu\text{m}$ for the inner and outer layers, respectively. Figure 1.7a shows the PXD ladder. The pixels are composed of DEPFET sensors. The DEPFET is a semiconductor detector concept that combines detection and amplification within one device. A cross section through the device is shown in Figure 1.7a. A p-channel MOSFET or junction field effect transistor (JFET) is integrated onto a silicon detector substrate, which becomes fully depleted by a sufficiently high negative voltage to a p⁺ contact on the back side. A potential minimum is formed by sideward depletion [17], which is shifted directly underneath the transistor channel at a depth of about $1\ \mu\text{m}$ by an additional phosphorus implantation underneath the ex-

ternal gate. Incident particles generate electron-hole pairs within the fully depleted bulk. While the holes drift to the back contact, electrons are accumulated in the potential minimum, called the “Internal Gate”. When the transistor is switched on, the electrons modulate the channel current. After being read out, the internal gate gets reset by the control circuitry and is thus ready for the next cycle. During the readout the sensitive area integrates all of the charge accumulated under the internal gate, thus the integration window is approximately equal to the total readout time.

The readout technique is based on a synchronous serial [3] frame sweep (rolling shutter mode). That is, pixels are read out in series (pixel by pixel). The pixel matrix is split into two 800×250 pixel frames that make up two half-ladders. Each frame is connected to a switcher ASIC that in turn routes four pixels in parallel to four DEPFET Current Digitizer ASICs (DCDBs) for digitization. These DCDB chips are followed by four DHP chips designed to transfer the data off the half-ladder. The entire PXD is composed of 40 half-ladders. The readout time necessary to complete a full half-ladder sweep is $\approx 20 \mu s$ (two full SuperKEKB orbits). This long integration time leads to an estimated occupancy of up to 3 % for the innermost layer. The expected PXD output data rate at the highest occupancy is approximately 58 Gbits/s with an event size of approximately 1 MB per event. The spatial resolution of the detector has been reported to be approximately $14.4 \mu m$ [3,18].

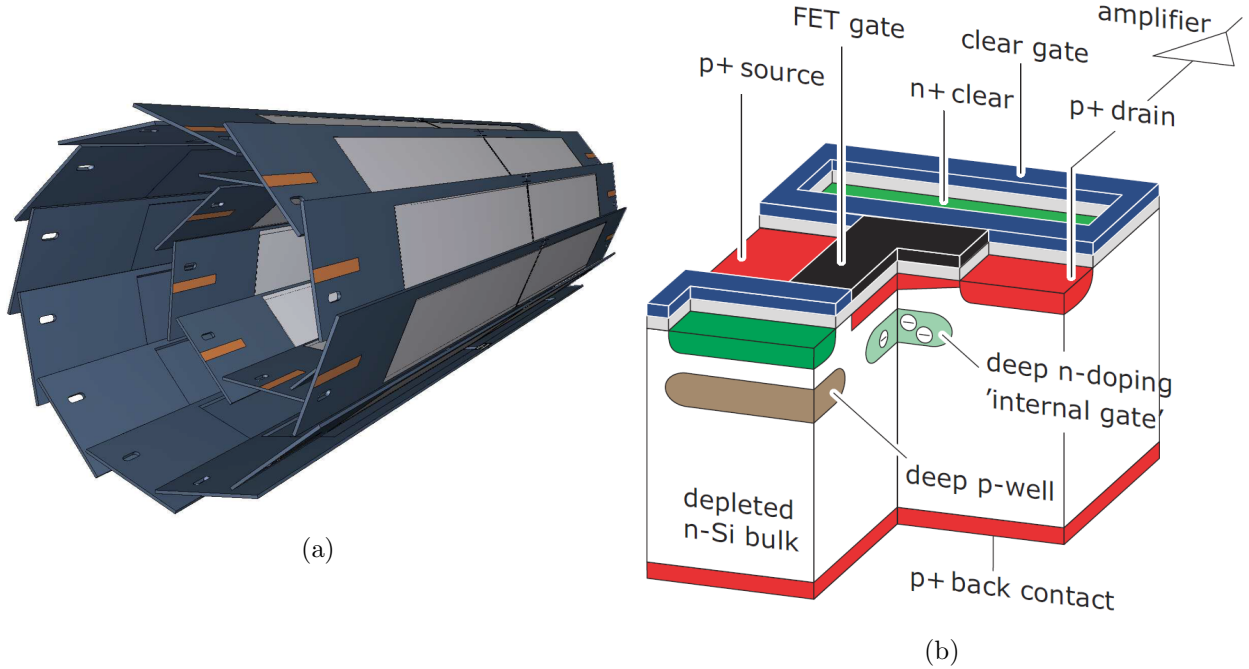


Figure 1.7: (a) Schematic view of the geometrical arrangement of the sensors for the PXD. The light grey surfaces are the sensitive DEPFET pixels thinned to $50 \mu m$ and covering the entire acceptance of the tracker system. The full length of the outer modules is 174 mm [5]. (b) Operating principle and structure of a DEPFET [6].

1.3.2 Summary of Vertex Detectors

Vertex detectors can be divided in two categories: pixels based and strip based. The major advantage of the pixel based vertex detectors is a higher number of channels and thus lower occupancy for the same hit rate densities compared with the strip-based vertex detectors. However, the disadvantages are a more complex architecture and readout, higher material budget, higher power consumption, and much higher output data rates, which present a considerable challenge for the processing electronics.

The high occupancy (3 %), high data rate (58 *Gbits/s*), and the 20 % dead time required during the injections of the Belle II PXD are significant drawbacks for the experiment. Furthermore, the PXD data rate dominates the overall data volume of the spectrometer as illustrated in Figure 1.8.

In the following chapter we present an alternative vertex detector architecture that combines the best of both worlds (pixel and stripline), thus offering a considerably lower occupancy as well as lower data rates at a comparable spatial resolution to the PXD, thus presenting itself as a potential candidate for a future upgrade of the PXD.

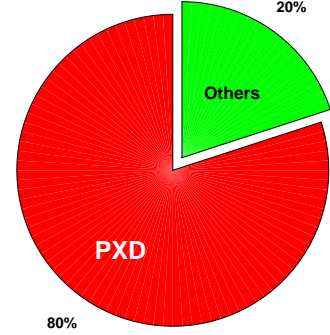


Figure 1.8: The PXD data volume in comparison with the rest of the spectrometer.

CHAPTER 2

TIMING VERTEX DETECTOR

In this chapter, we introduce a novel vertex detector architecture. Its design relies on an asynchronous digital pixel matrix in combination with a readout based on high precision time-of-flight measurement. Denoted the TVD, it consists of a binary pixel array, a transmission line for signal collection, and a readout ASIC. The main objectives of the TVD are:

- achieving a spatial resolution comparable to the existing Belle II vertex detector;
- reducing the occupancy by a factor of ten;
- decreasing the channel count by almost three orders of magnitude.

The subsequent sections introduce the TVD novel architecture, describe its principle of operation in detail, and explore the feasibility of each subpart as well as give simulation results on some of the key performance parameters, which are then used for direct comparison with the Belle II PXD. At the same time, design specifications are derived and summarized at the end of the chapter, thus laying the foundations for the research and development (R&D) of the TVD.

2.1 TVD Sensor Ladder Architecture

In this architecture, a pixel is composed of an active detection area coupled to an electronic circuit simultaneously serving as a driver and as a buffer. The pixel matrix is arranged in a row-column format, where all of the pixels in one row are connected in parallel along the transmission line. This forms one readout channel. Each readout channel is connected to a reflective termination on one end (far-end) and to the input of the readout ASIC on the other end. For the purpose of this study and for the ease of comparison, we adopted the pixel array dimensions matching those of the Belle II PXD sensor ladder [3, 19]. The dimensions of the active area per ladder ($A_{act./lad.}$) are $80\text{ mm} \times 12.5\text{ mm}$ or 10 cm^2 for the inner layer and $120\text{ mm} \times 12.5\text{ mm}$ or 15 cm^2 for the outer layer. Figure 2.1 shows a 3D model sketch of the proposed TVD ladder architecture.

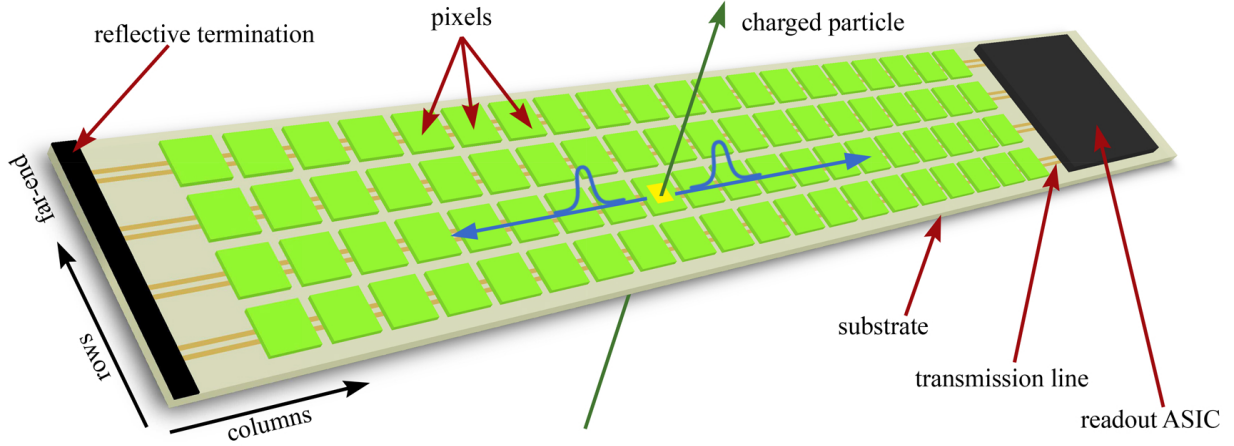


Figure 2.1: 3D sketch of the proposed TVD sensor ladder architecture.

A key advantage of this architecture is the ability to provide a scalable and, to some extent, large number of pixels while employing relatively few readout channels. Consequently, this leads to a significant reduction of the output data rate, power consumption, and material budget, possibly also reducing the active cooling requirements. Furthermore, this architecture allows the system design to be divided into three semi-independent subparts: the pixel, the transmission line, and the readout ASIC. The fourth subpart of the design is the interconnect technology that needs to be used to connect the first three subparts without degrading their performance. In the case of the pixel to transmission line connection, the use of TSVs is envisioned. The readout ASIC foresees the use of the flip chip technology with the die directly positioned on the sensor surface. The pitch of the ASIC inputs will match the pitch of the transmission line and hence the pixel array.

2.2 Principle of Operation

This section describes the principle of operation of the detector in detail, including the expected occupancy and spatial resolution. Subsequently, the target timing resolution is derived.

Each readout channel operates independently of all others. Pixels within one readout channel are operated in an asynchronous and autonomous mode. After detecting a traversing charged particle in the active area of the pixel, the in-pixel electronics responds by issuing a set-reset cycle, forcing the active area to discharge and the output buffer to produce a voltage pulse. Upon injection into the transmission line, the voltage pulse splits into a forward propagating pulse (towards the readout ASIC) and a backward propagating pulse (towards the far-end). The backward propagating pulse is reflected at the far-end and travels back towards the readout ASIC. The arrival time of both voltage pulses is measured by the readout ASIC and the pixel position is determined by employing time-of-flight (TOF) equations [20]. Figure 2.2 illustrates the whole process.

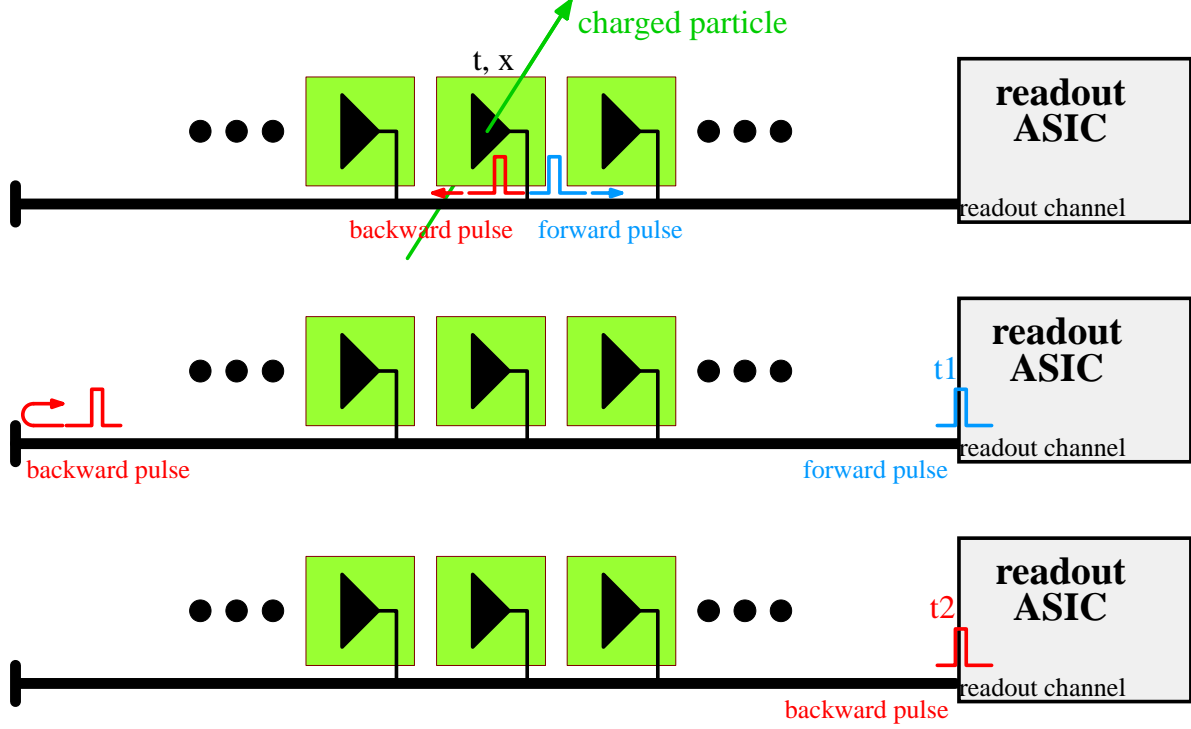


Figure 2.2: Voltage pulse generation and injection into the transmission line (top). Pulse propagation in the transmission line (center). Measurement of the pulse arrival time (bottom).

The occupancy of a given channel strip depends on the time necessary for the pulses to travel to the readout ASIC and clear the transmission line. In other words, the occupancy is directly proportional to the propagation delay between the two pulses. The spatial resolution depends on the time of arrival precision in the TOF equation. The spatial and time resolutions are related through the propagation speed as postulated by the wave equation.

Both, the propagation speed and the propagation delay, are directly dependent on the sensor substrate relative permittivity (ϵ_r). This dependency is shown in Figure 2.3. The permittivity range shown here considers substrates with low ϵ_r , from FR-4 [21], up to higher ϵ_r substrates like silicon. The overall delay between the two pulses depends on the fired pixel position on the sensor ladder. The largest delay between the two pulses occurs for the pixels positioned closest to the readout ASIC. This delay is used to define the worst case time window.

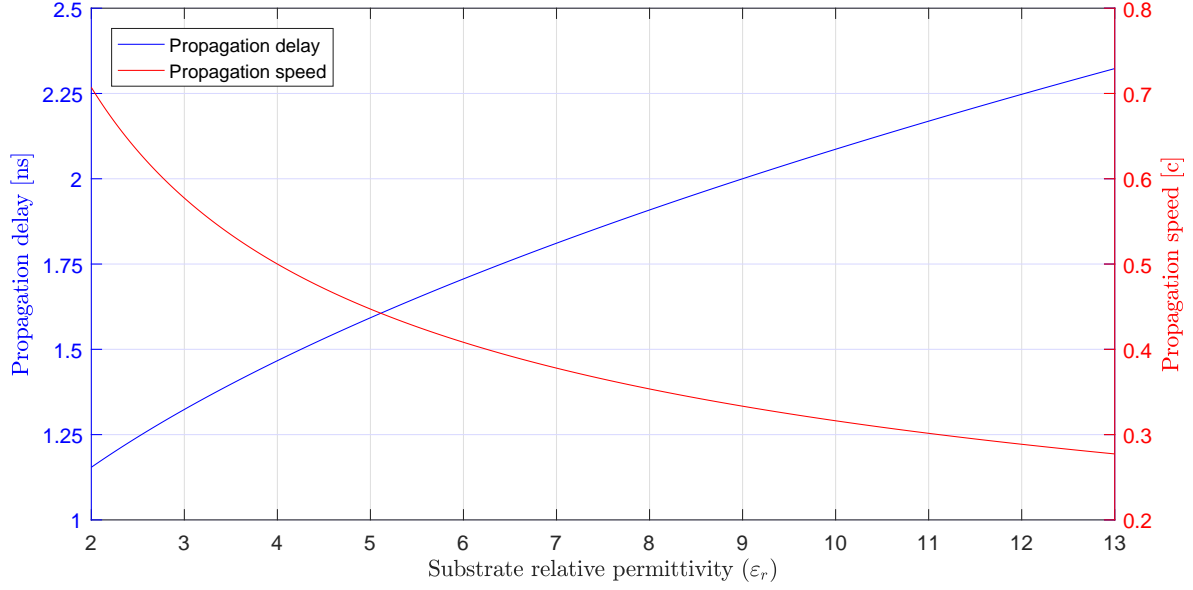


Figure 2.3: Propagation delay and propagation speed as functions of the substrate ϵ_r for pixels close to the readout ASIC with a sensor ladder length of 90 mm.

Propagation delay decreases with the decreasing ϵ_r , hence shortening the theoretical minimum time window. However, a decreasing ϵ_r increases the propagation speed, thus degrading the spatial resolution for a given timing resolution. The single point spatial resolution limit is set by the variance of the associated continuous random variable [22]. That is, the pixel longitudinal size divided by $\sqrt{12}$. Hence, the theoretical spatial resolution limit is approximately 14.4 μm . However, that would imply an almost zero root mean square (RMS) error of the timing resolution, which is unachievable in practice. In practice the overall spatial resolution (σ_{space}) is the quadrature sum of the spatial resolution derived from the timing resolution (σ_{time}) and the inherent single point spatial resolution limit, as shown in Equation 2.1. The overall spatial resolution dependence on the propagation speed and the timing resolution is plotted in Figure 2.4.

$$\sigma_{space} = \sqrt{(\sigma_{time} \cdot v_p)^2 + \left(\frac{50 \mu m}{\sqrt{12}}\right)^2} \quad (2.1)$$

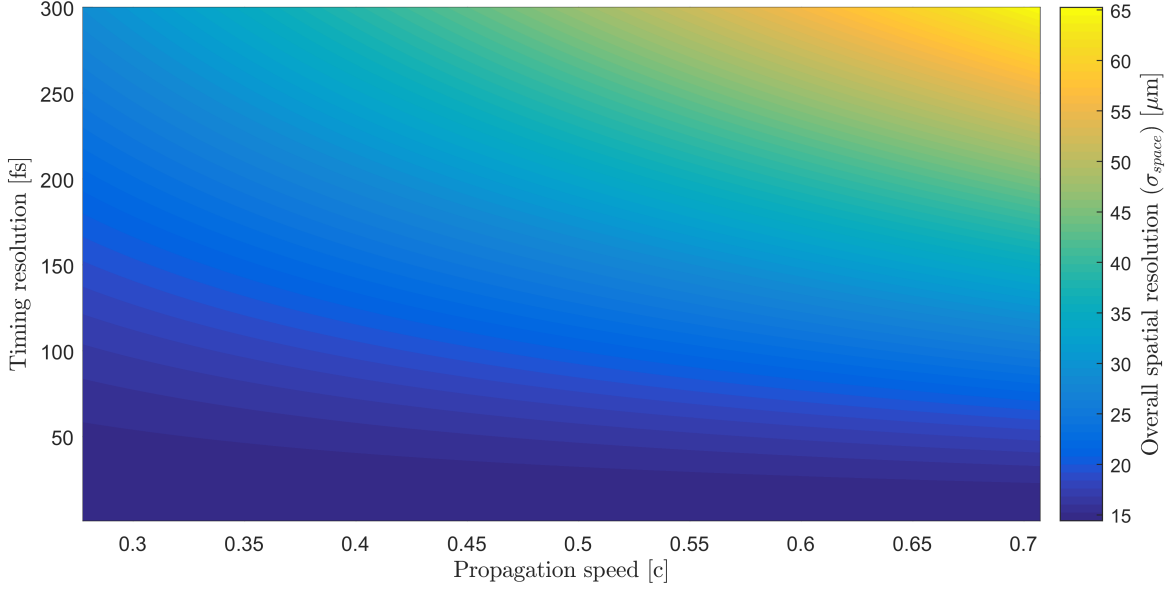


Figure 2.4: Spatial resolution as a function of propagation speed and timing resolution.

Considering the results shown in Figure 2.4, the timing resolution has to be kept in the range of 100 fs or less in order to get close to the overall target Belle II spatial resolution of approximately 10 μm [3]. A lower propagation speed, and thus higher ϵ_r , is more favorable in terms of timing resolution requirements.

The occupancy is evaluated for each readout channel independently. To avoid or reduce the effect of pileup, the maximum hit rate that the channel can tolerate is set by the largest propagation delay between the two pulses. For $\epsilon_r \approx 11.7$ this delay is approximately 2 ns . Therefore, the expectation value of the hit rate for 100 % occupancy is estimated to be on the order of $5 \cdot 10^8 \text{ Hits/channel} \cdot s$. Considering that there are 250 channels per ladder and the $A_{cat./lad.}$, the saturated hit rate density expectation value ($E[\max(\rho_{hit \text{ rate}})]$) is $1.25 \cdot 10^{10} \text{ Hits/cm}^2 s$. The Belle II PXD is composed of eight ladders in the innermost layer and twelve ladders in the outer layer, each with $4 \cdot 10^5$ pixels (1600×250) and a frame readout time of 20 μs per ladder. For these $8 \cdot 10^6$ pixels, the reported 3 % occupancy means $2.4 \cdot 10^5$ occupied pixels per event [18]. The $E[\max(\rho_{hit \text{ rate}})]$ for the Belle II PXD is thus $1.54 \cdot 10^9 \text{ Hits/cm}^2 s$. Table 2.1 summarizes the parameters and the calculation results for both, the Belle II PXD and the proposed TVD.

Table 2.1: Summary and comparison of the Belle II PXD and the TVD

<i>Parameter</i>	<i>Belle II PXD</i>	<i>Proposed TVD</i>
Inner layer pixel array dim.	80 mm × 12.5 mm	80 mm × 12.5 mm
Outer layer pixel array dim.	120 mm × 12.5 mm	120 mm × 12.5 mm
Inner layer $A_{act./lad}$	10 cm ²	10 cm ²
Outer layer $A_{act./lad}$	15 cm ²	15 cm ²
Number of ladders	20	20
Number of rows per ladder	250	250
Number of pixels in total	$8 \cdot 10^6$	$8 \cdot 10^6$
Number of readout channels	$8 \cdot 10^6$	5000
Integration time	20 μs	2 ns
$E[\max(\rho_{hit\ rate})]$	$1.54 \cdot 10^9\ Hits/cm^2s$	$1.25 \cdot 10^{10}\ Hits/cm^2s$
Occupancy at $4.62 \cdot 10^7\ Hits/cm^2s$	3%	0.37 %
Average number of occupied readout channels per event	$2.4 \cdot 10^5$	18.5

Figure 2.5 shows how the estimated occupancy increases as a function of hit rate density for both, the Belle II PXD and the proposed TVD. The dotted vertical and horizontal lines indicate the equivalent occupancy of the proposed TVD to 3 % occupancy reported for the Belle II PXD. In addition, the dashed vertical line indicates the occupancy of the proposed TVD with an added dark count rate of 1 MHz per channel taken into account. With that the occupancy increases from 0.37 % to 0.57 %. It is also possible to relate the occupancy and spatial resolution to the substrate permittivity. Figure 2.6 shows the occupancy and spatial resolution as functions of ϵ_r .

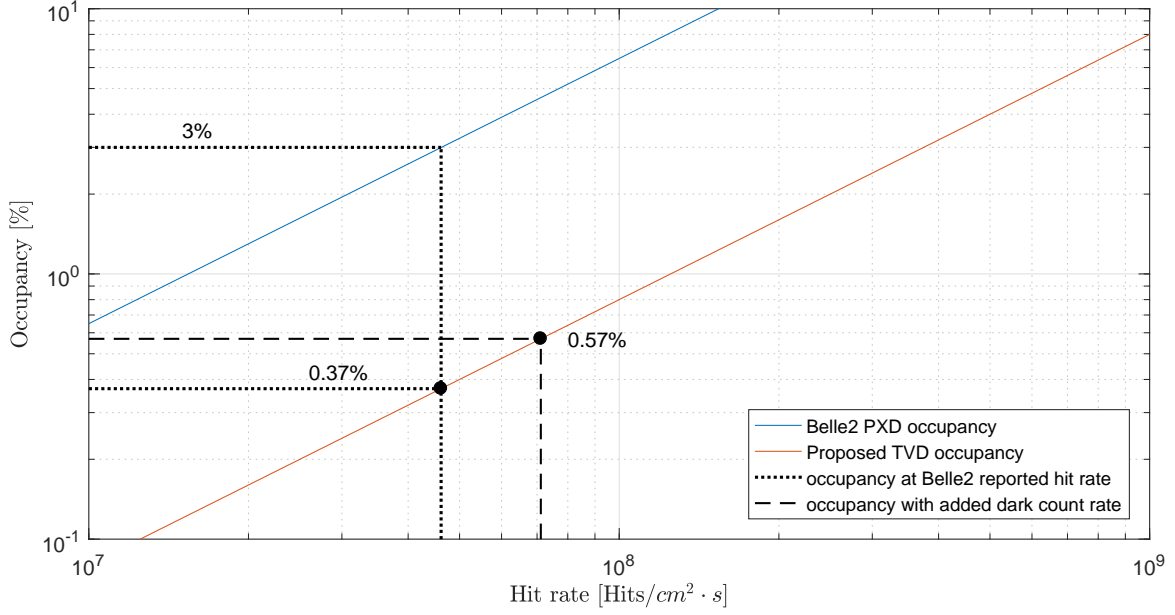


Figure 2.5: Occupancies of the Belle II PXD and the proposed TVD as functions of hit rate density.

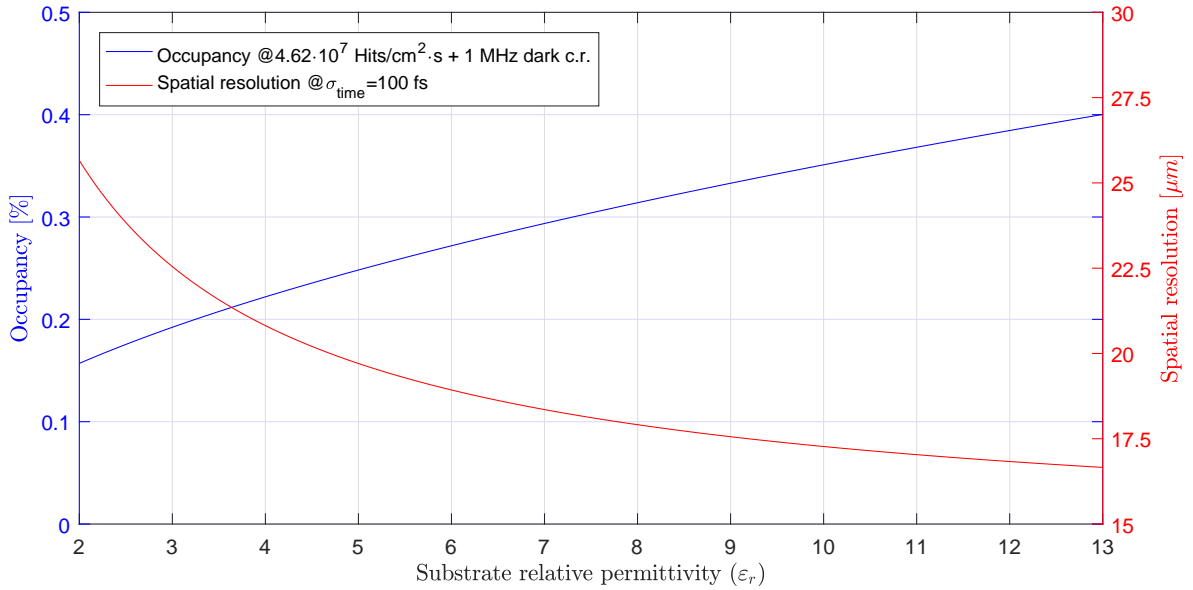


Figure 2.6: Occupancy at a $\rho_{hit\ rate} = 4.62 \cdot 10^5\ Hits/cm^2s$, 1 MHz dark count rate per channel, and spatial resolution at timing resolution of 100 fs as functions of ϵ_r .

Multiple particle hits are possible in a single channel, thus producing multiple signals on the transmission line. Figure 2.7 shows the probability of multiple hits occurring in one channel in a given time window. Most of the time there are no hits. The probability of a single hit is equal to the

occupancy, while the probability of having two or more hits in the same time window is $\approx 10^{-3} \%$ assuming the worst-case hit rate. These numbers have been obtained by using the Equation 2.2, which is the formula for the Poisson probability density function,

$$P(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}; \quad k = 0, 1, 2, 3, \dots, \infty \quad (2.2)$$

where k is the number of hits per time window and λ is the average number of hits per channel per time window.

Considering this very low number, we expect to simply identify a pileup in the superimposed waveforms and reject such hits as an efficiency loss.

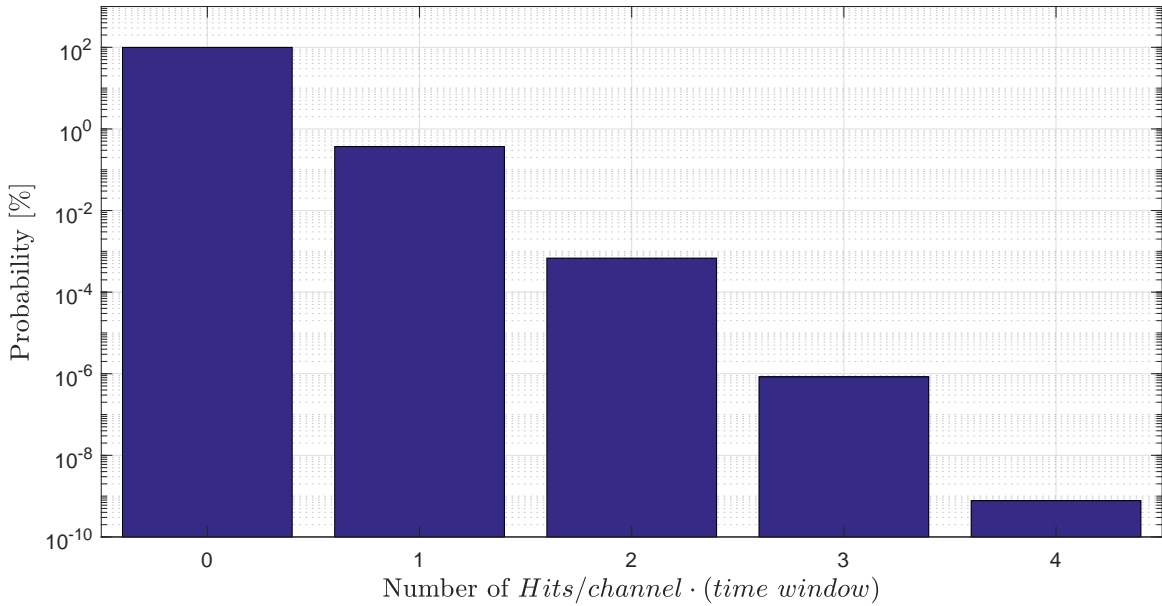


Figure 2.7: Probability of multiple event hits per channel per time window at the worst-case hit rate.

Furthermore, given the Belle II acceptance angle ($\alpha_{acc.}$) of 17° in the forward direction and 30° in the backward direction [3], one charged particle can hit two adjacent pixels at the same time, thus inducing two pulses, where one is the proper one and the other is the redundant, thus contributing to the overall occupancy (OC). The OC resulting from the redundant pulses has been calculated using Equation 2.3.

$$OC = OC_{\rho_{hit \ rate}} + \rho_{hit \ rate} \cdot \sum_n^{1600} A_{acc_n}(\alpha_{acc}) \cdot tw \quad (2.3)$$

Where tw is the time window of the sensor (2 ns) and A_{acc} is the acceptance area of each pixel, which is dependent on the acceptance angle and has been calculated by using Equation 2.4.

$$A_{acc} = W_{pix} \cdot \frac{H_{pix}}{\tan(\alpha_{acc.})} \quad (2.4)$$

Where W_{pix} is the pixel width ($50 \mu m$) and H_{pix} is the thickness of the pixel active area. Figures 2.8a and 2.8b show the acceptance area over the pixel total area ratio as a function of the pixel active area along the sensor ladder and the occupancy as a function of the pixel active area thickness, respectively. The occupancy does not increase significantly as the increases in thickness.

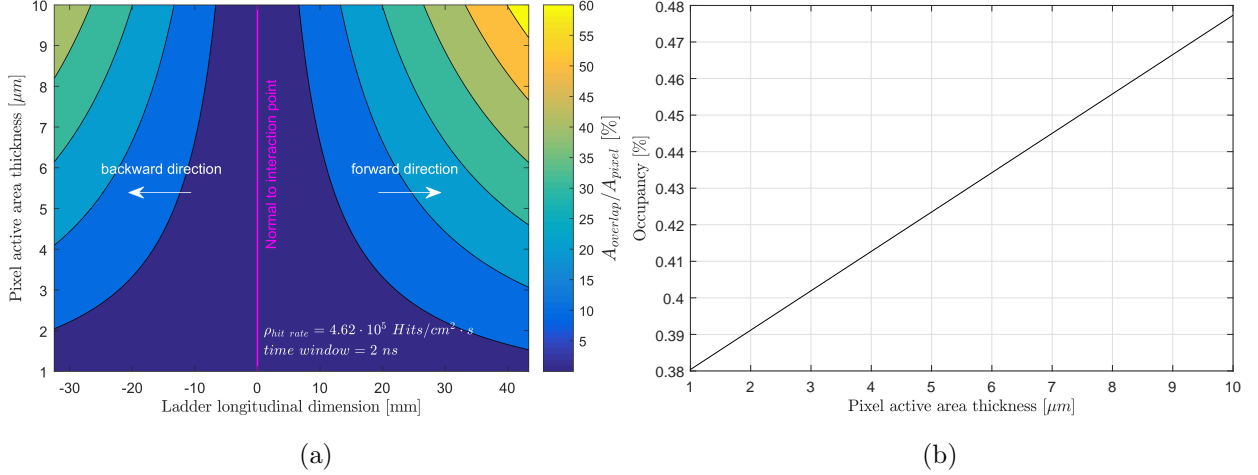


Figure 2.8: (a) Acceptance area over pixel total area ratio as a function of the pixel active area along the sensor ladder. (b) Overall increase in occupancy as a function of pixel active area thickness.

The data rate coming out of the TVD can be estimated by considering the fact that the maximum L1 trigger rate in the Belle II system is approximately 30 kHz with a fixed latency of roughly $5 \mu s$ [3]. With an average occupancy of 0.57% , the expected hit rate is approximately $37 \text{ Hits/event}/(time \text{ window})$. Furthermore, assuming a L1 trigger timing uncertainty of approximately 20 ns , ten time windows per hit need to be converted. Each time window requires a certain number of samples depending on the extraction algorithm used. A more complex curve fitting algorithm, which will be discussed later, requires 28 out of 40 samples per time window. Each sample can be encoded with 32 bits (twelve for amplitude, twelve for time stamp and eight for channel number). Equation 2.5 is used to calculate the average number of hits per event.

$$N_{Hits/event} = N_{channels} \cdot occupancy \cdot N_{search \text{ windows}} \quad (2.5)$$

Equation 2.6 is used to calculate the number of bits required per event.

$$N_{bits/event} = N_{Hits/event} \cdot N_{samples/Hit} \cdot N_{bits/sample} \quad (2.6)$$

Equation 2.7 is used to calculate the minimum data rate required in order to transfer all of the

data out of the proposed TVD.

$$Data\ rate = N_{bits/event} \cdot f_{trigger} \quad (2.7)$$

Table 2.2 summarizes the values used to calculate the data rate.

Table 2.2: Parametric values used in calculating the data rates for two different timing extraction algorithms with different numbers of waveform samples.

<i>Parameter</i>	$N_{samples/Hit} = 28$	$N_{samples/Hit} = 4$
$N_{channels}$	5000	
OC	0.57 %	
$N_{search\ windows}$	10	
$N_{bits/sample}$	32	
$f_{trigger}$	30 <i>kHz</i>	
Total TVD data rate	$\approx 10\ Gbits/s$	$\approx 1.4\ Gbits/s$

The aggregate data rate for 28 samples is approximately 10 *Gbits/s* or 500 *Mbits/s* per ladder. This represents a reduction factor of 5.8 compared to the Belle II PXD rates [3]. This number can vary depending on the number of samples required and the timing uncertainty of the trigger. For example, if a simple linear extraction algorithm is sufficient, only four samples per time window are needed hence reducing the total data rate to approximately 1.4 *Gbits/s* or 70 *Mbits/s* per ladder.

Having independent and asynchronous channels, the TVD offers the ability to differentiate between simultaneous signals occurring in different channels. However, the time resolution between channels also depends on the jitter of the pixel; thus, a timing resolution no better than a few picoseconds is expected. To avoid confusion, it should be noted that the sensor jitter does not affect the differential time measurement in the ASIC, because the original pulse splits passively after the pixel output buffer.

2.3 Feasibility Study, Requirements, Results, and Technical Challenges

2.3.1 Design Requirements

Different timing techniques, including single-threshold, constant fraction, multiple threshold, and waveform sampling, have been evaluated. It can be shown that the waveform sampling technique yields the best timing resolution performance [23]. In this work, we use this technique and the

associated assumptions as a starting point to derive the necessary design parameters to achieve femtosecond timing resolution.

The waveform sampling technique employs direct storage and digitization of the pulse waveform, interpolation of the sampled points, and comparison of the interpolated data with a fixed threshold value. As a result, a timing resolution better than the sampling interval can be achieved. Furthermore, it can be shown that, after appropriate calibration, a good estimation of the achievable theoretical minimum timing resolution is proportional to noise (Δu) and inversely proportional to the square root of the product of sampling frequency ($f_{SAMP.}$) and analog bandwidth (BW) [24] as shown in Equation 2.8. U is the signal amplitude.

$$\sigma_t = \frac{\Delta u}{U} \cdot \frac{1}{\sqrt{3f_{SAMP.} \cdot BW}} \quad (2.8)$$

Timing resolution parameter sweeps in terms of bandwidth, noise, and sampling frequency, while assuming full signal amplitude ($1 V_{PP}$), are shown in Figure 2.9.

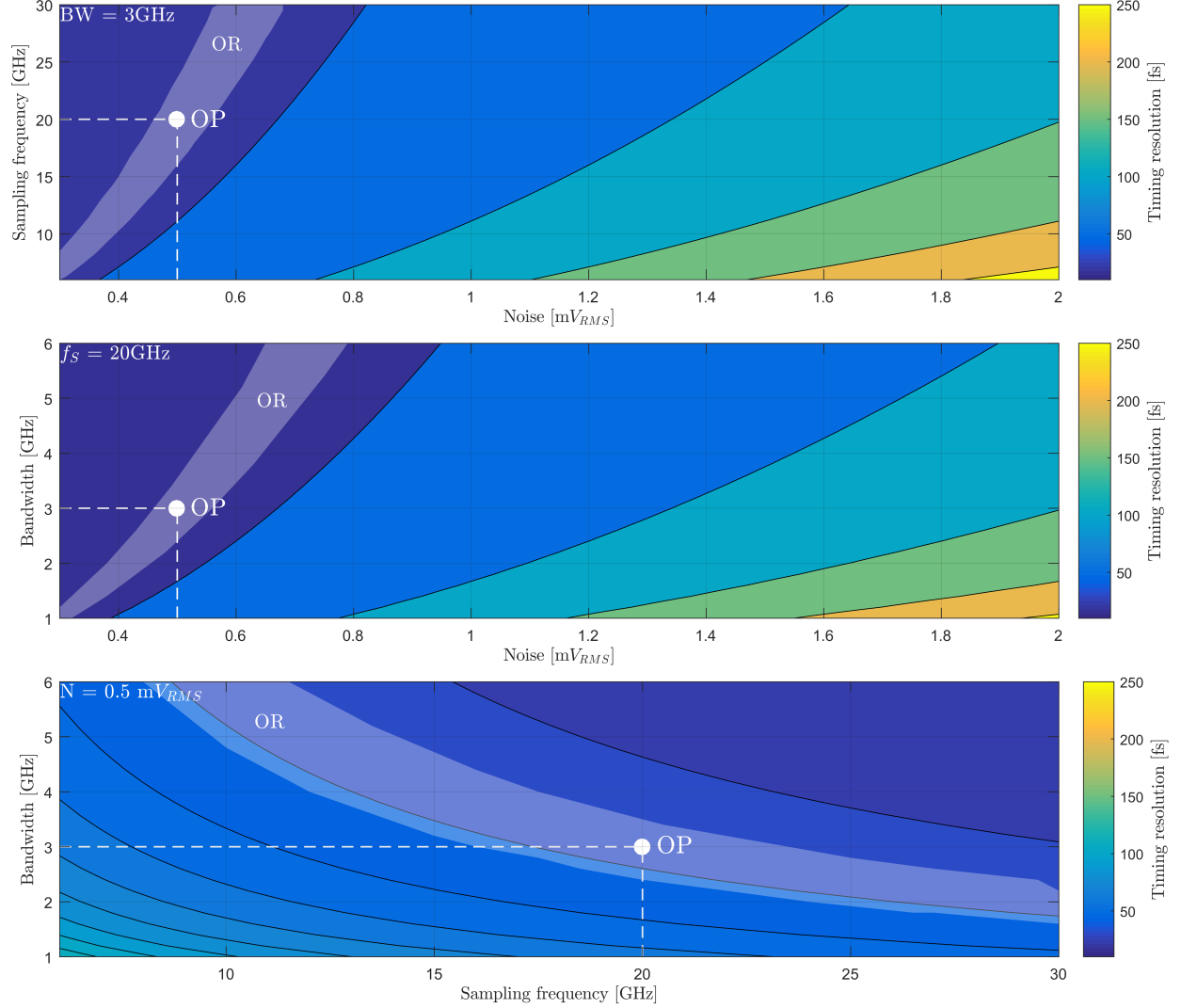


Figure 2.9: Timing resolution parameter sweeps in terms of f_{SAMP} and Δu (top), BW and Δu (center), and f_{SAMP} and BW (bottom) with clear indications of the desired operating region (lightly shaded area) and the nominal operating points (OP).

Considering a timing resolution target of 100 fs or less and a $\pm 10\%$ tolerance interval, viable operating regions (ORs) have been heuristically determined and indicated as lightly shaded regions on the graphs in Figure 2.9. In order to maximize the stability of the operating point (OP), the exact parameter values have been chosen so that the OP lies in the approximate center of all three operating regions. These values are summarized in Table 2.3.

As a result, the extrapolated theoretical timing resolution is estimated to be approximately 38 fs, which translates to an overall spatial resolution of approximately 15 μm . These requirements have been used to determine the initial design specifications, identify some of the technical

challenges, and evaluate the feasibility of each TVD sensor subpart. Figure 2.10 shows the relations between the design requirements and the TVD subparts.

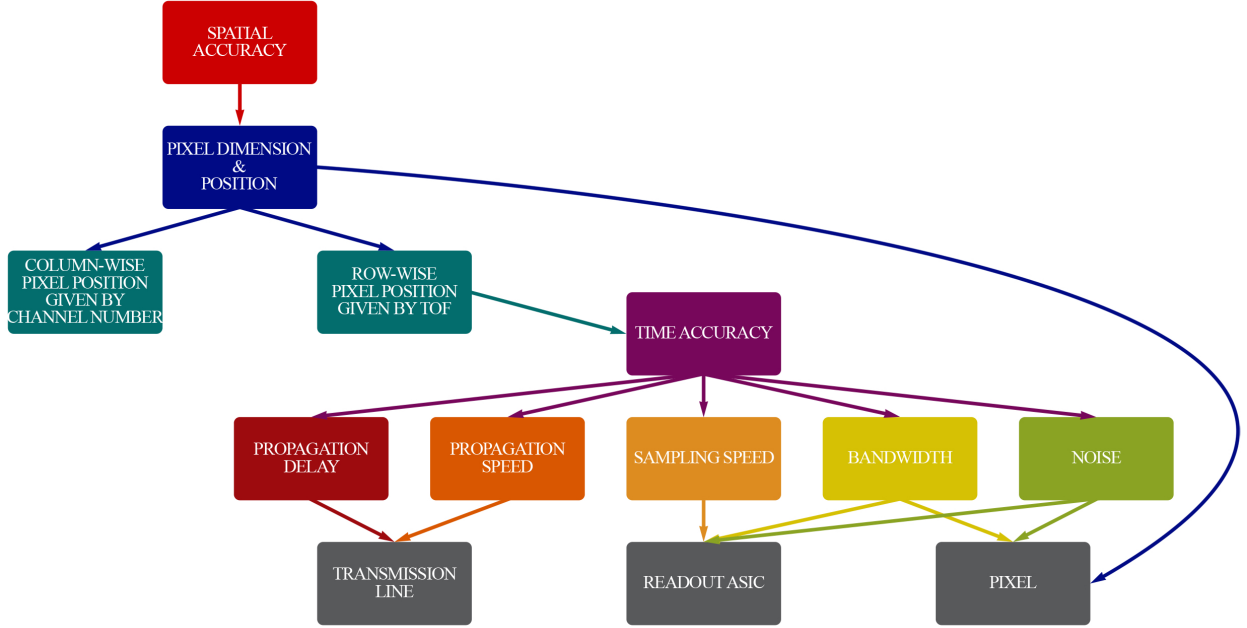


Figure 2.10: Direct relations between the design requirements and the corresponding subparts.

Table 2.3: A summary of TVD sensor baseline design parameters.

<i>Parameter</i>	<i>Value</i>
Noise	0.5 mV_{RMS}
Sampling frequency	20 GHz
Analog bandwidth	3 GHz

2.3.2 Transmission Line

The ability to fully distinguish the two pulses depends on their full width at half maximum (FWHM) and their separation in the time domain. Figure 2.11 shows the expected pulse shape. The rising and falling edges have been approximated with a Gaussian function.

The bandwidth requirement imposes rise and fall times of around 120 ps . In addition, a flat-top of 50 ps has been assumed in order to obtain some margin for error. As a result, the pulse FWHM is around 215 ps . The minimum peak distance of the pulses has been determined so that the overlap of the pulses does not cause the superposition of waveforms to exceed 10% of the pulse

amplitude. This coincides with the lower voltage threshold used to determine the rise time. As a result, the minimum peak separation ($\Delta\tau_{pp}$) has been found to be around 400 ps. Consequently, the minimum added transmission line length between the farthest pixel from the ASIC input to the reflective far-end termination ($\Delta l_{PtoT.min}$) has to be equal to or greater than half of the minimum peak separation in spatial domain (Δl_{pp}). In this case, the distance is approximately 18 mm for a substrate $\epsilon_r = 11.7$. The delay of the transmission line and thus the time window is increased by 400 ps, which increases the worst-case occupancy from 0.57 % to 0.68 %.

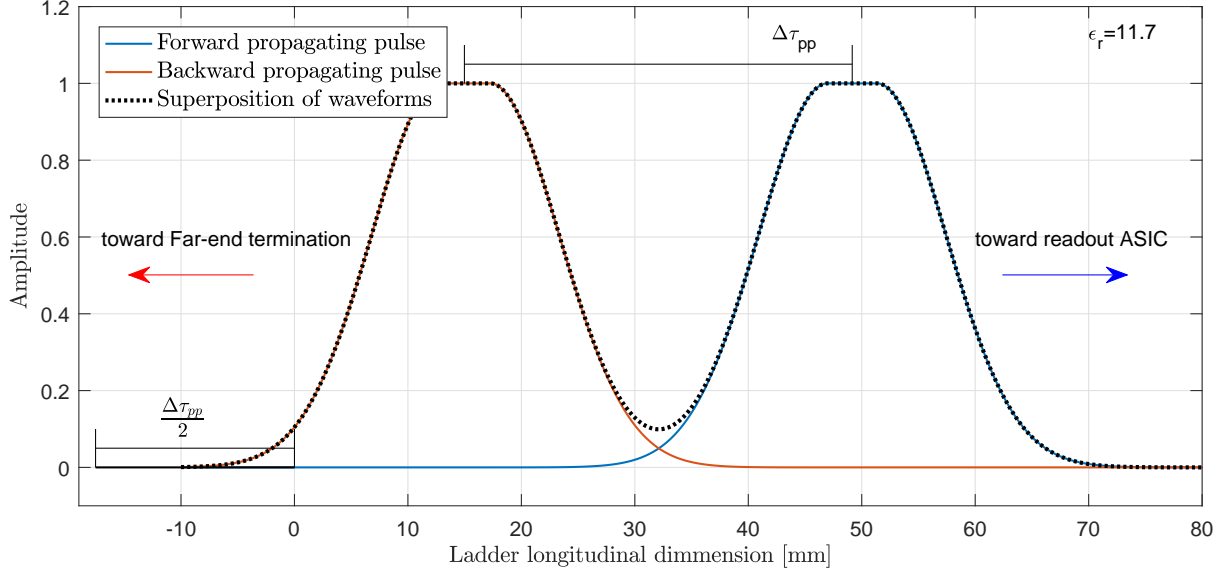


Figure 2.11: Voltage pulse approximation on the transmission line.

High permittivity substrates such as Gallium Arsenide (GaAs) ($\epsilon_r = 13.1$) [25], High-resistivity Silicon (HiresSi) ($\epsilon_r = 11.7$) [26, 27] or, alternatively, a printed circuit board (PCB) substrate ($\epsilon_r = 12.2$) [28] feature high linearity, low radio frequency (RF) losses, and high resistivity. The loss tangent ($\tan \delta$) at 10 GHz for these materials ranges from 0.0016 for GaAs and 0.0019 for the PCB substrate, to around 0.004 for the HiresSi. Simulation results of the transmission line indicate that losses are dominated by the limited conductivity of the traces, while the losses associated with the $\tan \delta$ are negligible. Figure 2.12 shows the cross section of the simulated transmission line geometry [29]. The transmission line loss simulation was performed using Keysight ADS [30], the results are shown in Figure 2.13a.

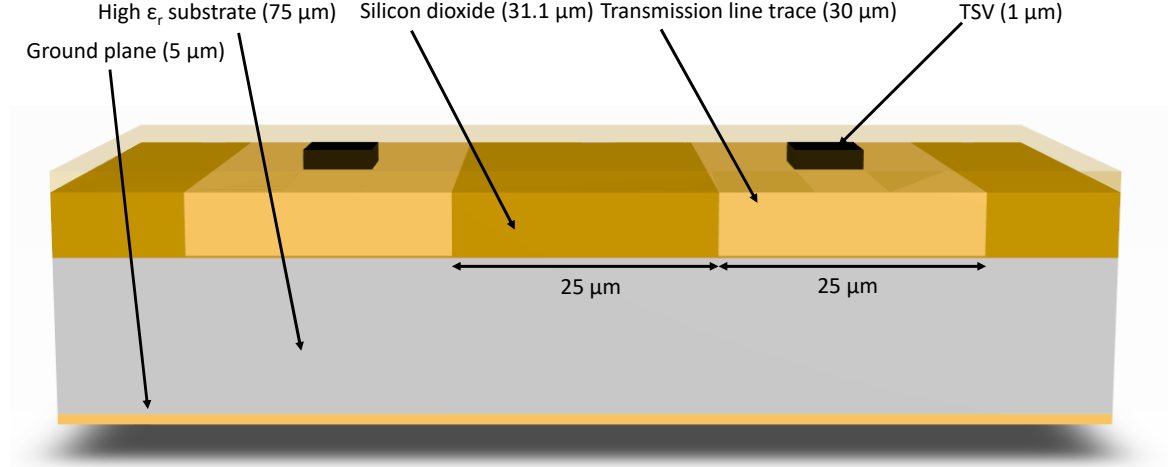


Figure 2.12: Cross section of the transmission line geometry with the vertical dimensions given in brackets.

For the longest signal travel path length of 19.6 cm , the loss of the transmission line is around 5 dB at 3 GHz . This lowers the signal-to-noise ratio (SNR) as well as the analog system bandwidth. Consequently, the timing resolution in the chosen OP effectively degrades from 38 fs to 66 fs . Figure 2.13b shows a calculation of the timing resolution as a function of signal power at a fixed bandwidth of 1.5 GHz . In order to maintain a timing resolution around 100 fs , the maximum signal attenuation should not exceed 5.5 dB .

Another important consideration is the phase drift. A measurement using a low bandwidth phase detector based on a saturated mixer principle [31] reveals a temperature dependence of the phase in the range of $0.7\text{ fs}/^{\circ}\text{C} \cdot \text{mm}$ in 0.3048 m long coaxial RF cables. Figure 2.14 shows the results of the measurement with a clear correlation between the phase drift and temperature. Similar effects are expected in the transmission line and ASIC. Precise temperature monitoring in combination with a feed-forward algorithm will be required to counter these effects. Furthermore, the ASIC power consumption should be kept constant through the use of switchable shunt resistors in order to ensure that once the detector reaches thermal equilibrium, the temperature change will be reasonably small.

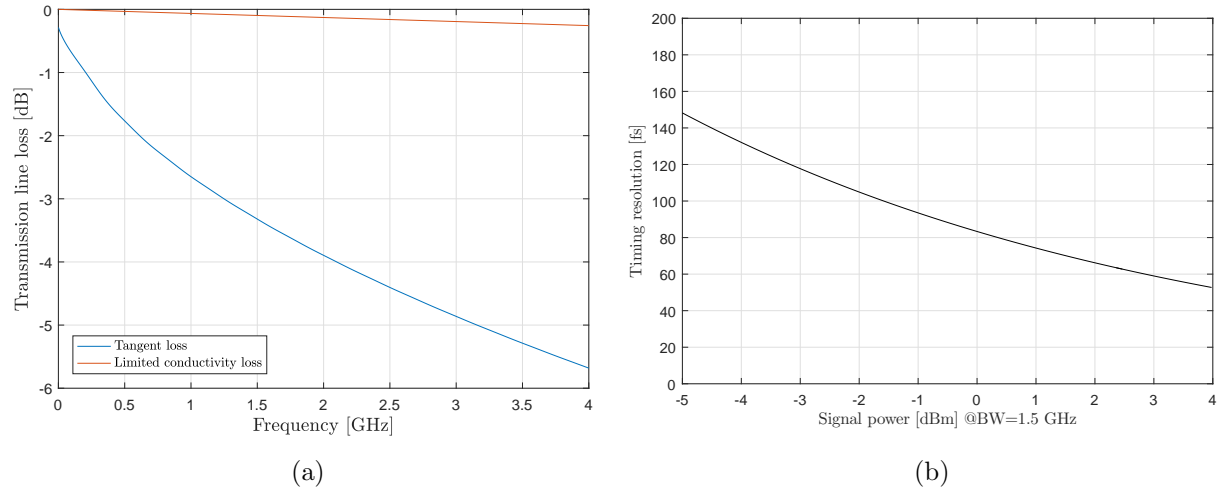


Figure 2.13: (a) Transmission line loss for the worst-case signal travel path of 19.6 cm. (b) Timing resolution as a function of signal power.

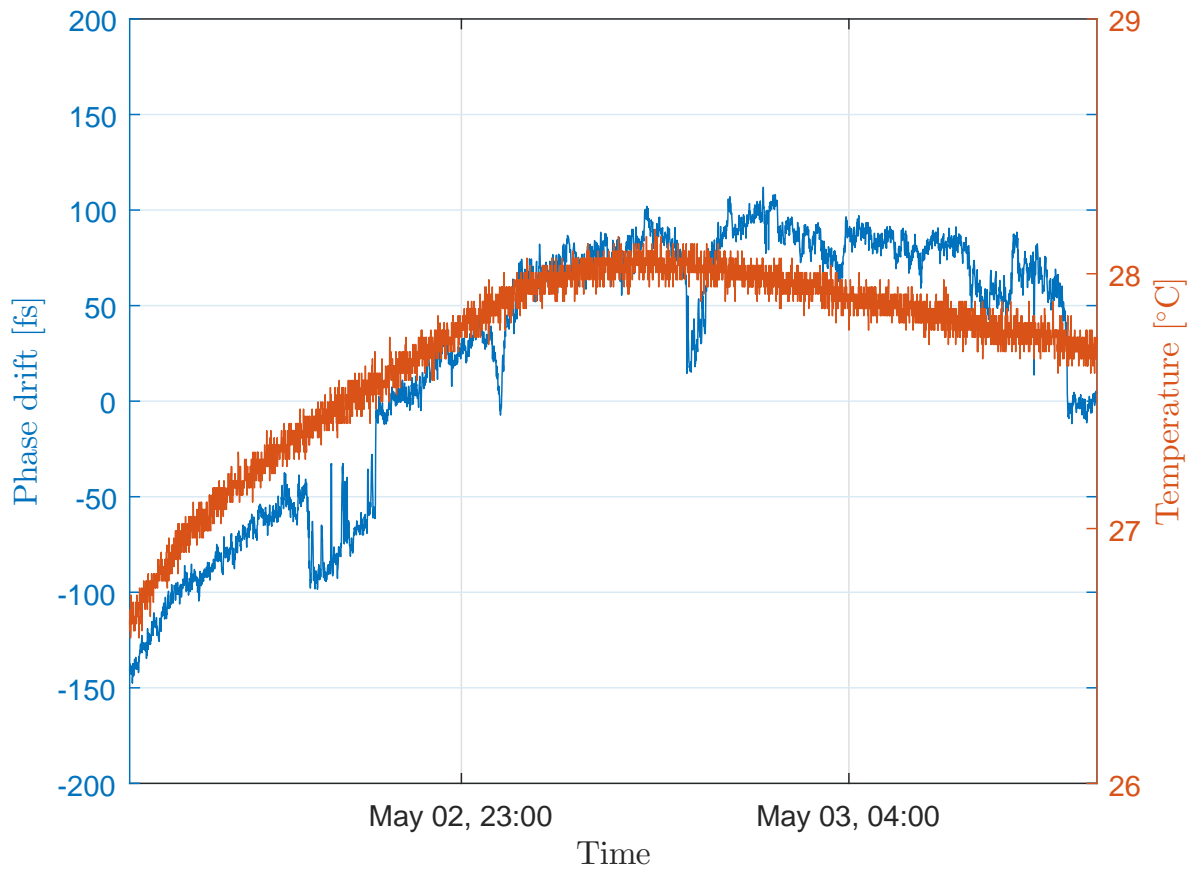


Figure 2.14: Measurement results of phase drift in cables over time with clear temperature dependence.

Every ASIC requires an integrated pulse generator in order to generate a pilot calibration pulse and inject it onto the transmission line. Using this pulse signal enables us to correct for errors arising from fabrication tolerances of the transmission line, mechanical strains due to temperature variation, and channel to channel differences of the ASIC.

Far-End Transmission Line Termination

The far end termination is needed to perfectly reflect the backward propagating pulse.

One option is to simply have the transmission line end with an open circuit, thus ensuring the boundary condition for total reflection. This option has the advantage to preserve the phase, noise, and jitter characteristics of the pulse. However, the pulse has to propagate all the way back to the ASIC, thus subject to more attenuation.

The second option is to build an ASIC with an integrated low noise and low phase noise amplifier as well as a directional coupler that can amplify the pulse amplitude. Figure 2.15 shows the simplified block schematic of one channel termination. With a proper voltage and current stabilization, a very low phase noise amplifier can be built with added jitter of around 1 fs . The obvious drawbacks are the additional power consumption, material budget, and group delay.

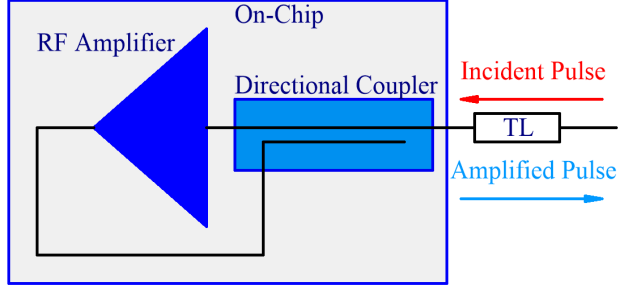


Figure 2.15: Single channel block diagram of the far-end termination ASIC with directional coupling and amplification.

2.3.3 The Pixel

To generate the required voltage pulses, a fast active detection element, along with fast and low noise buffer electronics is needed. Considering that energy resolution is not a requirement, technologies such as silicon photomultiplier (SiPM)/geiger-mode avalanche photodiodes (APD) [32], low gain avalanche photodiodes (LGAD) [33], 3D sensors [34, 35], or a combination of these could be a viable solution for the detection element of the Pixel. While SiPM structures were originally meant for photon detection, detection of charged particles is also possible [36, 37]. In recent years, significant progress has been achieved in timing response [38, 39], integration, and dark count rate mitigation [40]. The timing performances shown in this work are already sufficient for this application. Considering the expected occupancy of the detector, the mitigation of the dark counts is definitely a welcome feature, but not essential. And, in fact, resilience to higher rates will be important to cope with the expected increase in higher dark count rates (DCRs) due to neutron

damage. Figure 2.16a shows the simplified block diagram of the pixel, where coincidence logic is used to suppress the anticipated dark count accidental rate.

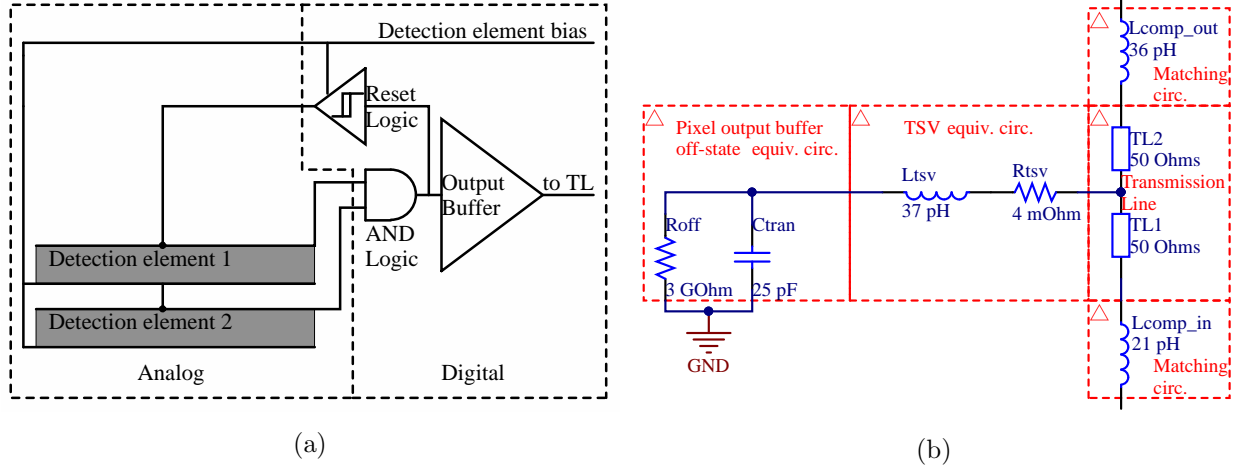


Figure 2.16: (a) Simplified block diagram of a single pixel cell. (b) Simplified equivalent circuit of the pixel driver, TSV, transmission line segment, and matching circuit.

The previously assumed dark count rate of 1 MHz per channel can be used to estimate the DCR per pixel. With 1600 pixels per channel, the tolerable DCR is reduced to 625 Hz . However, using the coincidence logic, the tolerable DCR can be increased significantly. Equation 2.9 is used for this estimate.

$$DCR_{w_coin.} = \sqrt{\frac{DCR_{w/o_coin.}}{\Delta t}} \quad (2.9)$$

In the cases where the coincidence time window (Δt) is assumed to be 2 ns , the $DCR_{w_coin.} \approx 560 \text{ kHz}$. Some reported DCR values for 130 nm CMOS technology based pixels are on the order of 25 kHz [41].

Recent results using thin devices in low-gain operation mode (so-called LGADs) have demonstrated intrinsically sufficient timing, since the requirements on the pixel firing time are decoupled from the spatial determination. To reduce collection time and further improve timing resolution, 3D electrode techniques also look very promising. Of particular interest is perhaps combining multiple, very thin layers with gain (either low or Geiger) into a coincidence. Such a sensor development is in longer term, though mentioned here as a herald that such a pixellated array with adequate performance may plausibly become available in the future.

Assuming a proper detection element, the in-pixel electronics has been modeled and its effect on the transmission line has been estimated. The off-state pixel output buffer parasitic capacitance, leakage resistance, and the TSV parasitic inductance and resistance [42] exert a distributed load on the transmission line, which causes signal attenuation and consequently degradation of the SNR.

In order to mitigate this attenuation, the output buffer has been matched to the transmission line. The current pixel output buffer circuit is being designed in a 130 nm CMOS process with low threshold level transistors. The simplified circuit used to model the pixel loading of the transmission line is shown in Figure 2.16b. The predicted output buffer off-state capacitance is in the range of 25 fF and the off-state resistance is 3 GΩ. The insertion loss results are shown in Figure 2.17b. The predicted pixel insertion loss as a function of the number of connected pixels has been observed to roughly follow a linear fit. The simulated overall insertion loss of the transmission line with 1600 pixels connected has been found to be approximately 5.3 dB. Figure 2.17b shows the simulated insertion loss of the transmission line with 1600 pixels as a function of frequency.

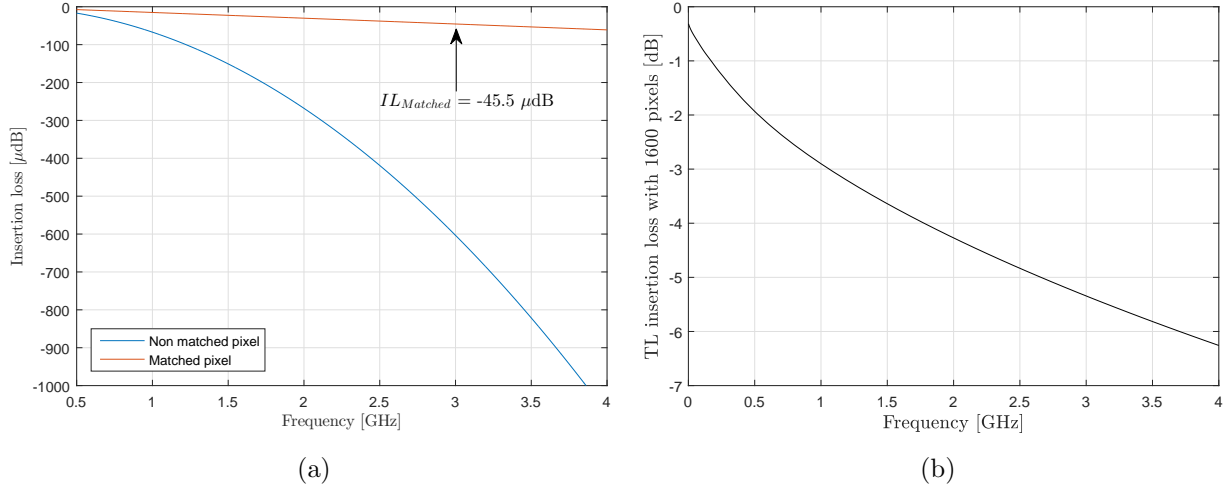


Figure 2.17: (a) Non-matched and matched single pixel insertion loss as a function of frequency. (b) Simulation results of a transmission line loaded with 1600 pixels.

The overall power consumption of the pixel is divided into dynamic and static components. The dynamic power consumption is dominated by the pixel output buffer when active, thus it is largely dependent on the hit rate. The static power consumption is composed of the quiescent current for the biasing of the comparator and the set-rest logic. With an occupancy of 0.68 %, the power density of the TVD pixel array is approximately 1.95 W/cm².

2.3.4 The Readout ASIC

Based upon the previous ASIC experience, an architecture using switched capacitor arrays (SCAs) has been chosen. Sampling speeds of up to 17 GS/s with bandwidths over 1.5 GHz and power consumptions on the order of 10 mW per channel have been demonstrated [43–45]. A major advantage of using this architecture is that the high speeds required for fast sequential sampling are limited to the SCA block. The subsequent transfer and digitization of the samples can be done in parallel at slower rates if necessary. Dead time during the transfer of the samples from

the SCA block can be mitigated by employing an interleaved sampling scheme with two or more SCA blocks. That is, when one SCA block is sampling, the other one is transferring and vice-versa. Consequently, the transfer rate is equal to the product of the sampling interval and the number of sampling cells in the SCA block. Considering an SCA block of 64 sampling cells, the transfer rate is reduced to 312.5 MHz , hence considerably relaxing the requirements of the subsequent stages.

The readout ASIC has been denoted as the **RFpix** and its simplified functional block diagram is shown in Figure 2.18. The RFpix has been divided into four circuits:

- Sample & Hold (S&H) circuit (described in detail in chapter five);
- Analog Storage and Digitization circuit (described in detail in chapter five);
- Data Transfer circuit (not covered in this work);
- Timing & Control circuit (see subsection below).

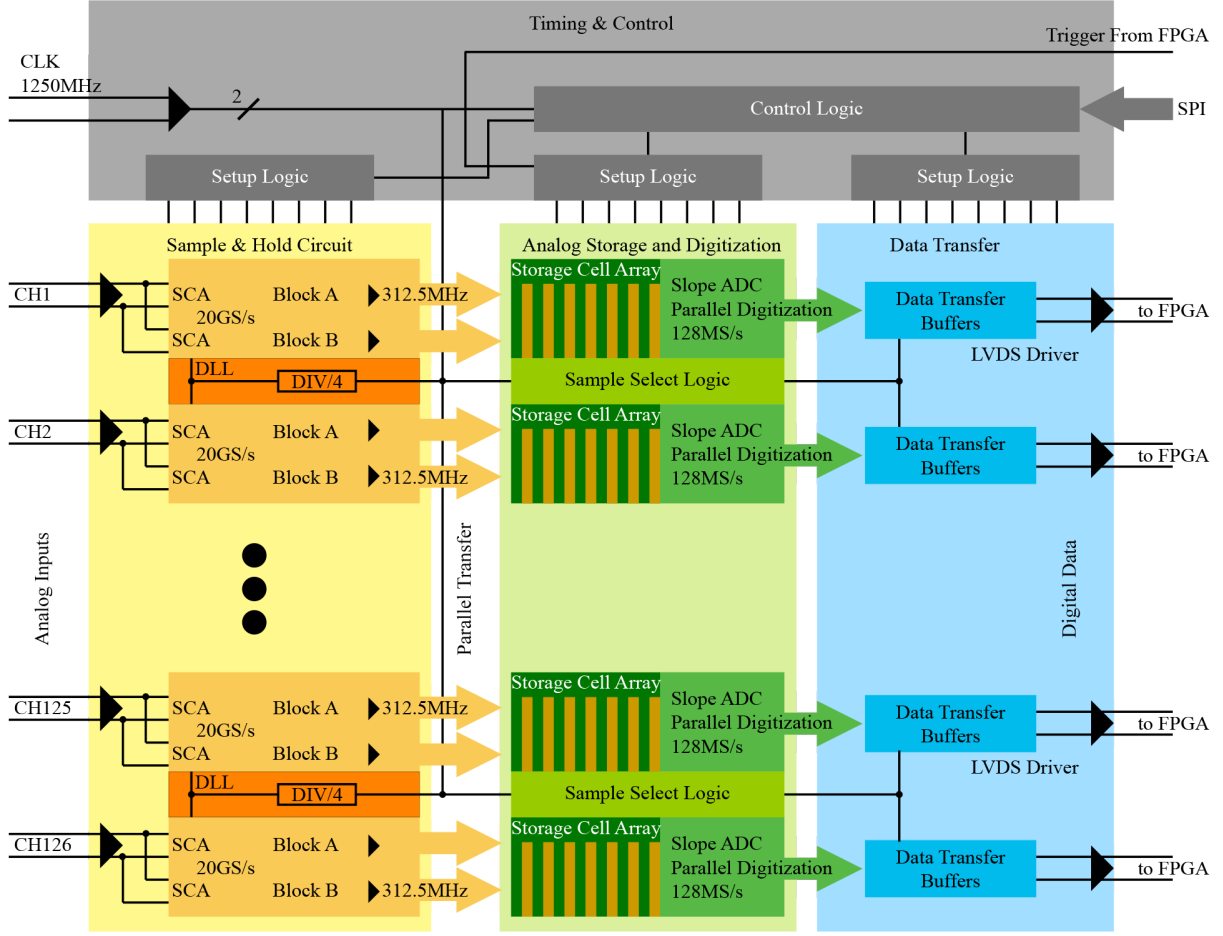


Figure 2.18: RFPix readout ASIC functional block diagram.

Timing & Control Circuit

The timing and control circuit is distributed through out the chip. The most complex part of the circuit is the clock distribution and generation network (CDGN). Jitter corrupts the timing resolution of the sampled signal through the sampling strobes. The added jitter of these strobe signals depends on the performance of the time base generator and the clock source.

In terms of sources, commercial crystal oscillator (XO) sources can provide jitter levels in the range of 10 fs to 20 fs [46]. If a worst case source jitter of 20 fs is taken into account, the jitter margin to reach the targeted 100 fs of timing resolution is on the order of 98 fs . In addition, considering the differential nature of the measurement the CDGN is expected to be the dominant term in the overall jitter of the system.

The CDGN is composed of:

- Input differential buffer that ensures the signal integrity of the input clock;
- A differential transmission line that fans out the buffered clock to all of the channels;
- Each pair of channels have an integrated differential two-level DLL that generates the sampling strobes for the sampling cells with low added jitter and fast rise times;
- Analog storage array timing logic;
- Output data transfer clock generator;

The control circuit is composed mainly of registers that control digital to analog converters (DACs) in order to set proper bias values for various circuits (digital and analog switches). The registers are loaded through a dedicated serial peripheral interface (SPI) that can be programmed from an external micro controller, another ASIC, or an field programmable gate array (FPGA), as is the case of the first prototype.

Estimated RFpix Power Consumption

Due to the low trigger rate of the application, the power consumption is dominated by the static and dynamic power consumption of the sampling-to-storage array buffers. Current estimates place this consumption at 23 mW per channel. The remaining power draw is attributed to the storage cell comparators (1.5 mW per channel), the ramp generator (3 mW per chip), the Wilkinson ADC counter (2.4 μW per chip), and the delay-lock loops (DLLs) (3.6 μW per channel). Considering 128 channels per chip, the total power consumption per channel is roughly 24.6 mW .

2.3.5 Initial TVD Subpart Design Specifications and Summary

The design requirements identified in the previous sections have been used to form a set of initial design specifications. These represent a concrete foundation for the current and future work and are summarized in Table 2.4.

Table 2.4: Baseline TVD design specifications.

<i>Subpart</i>	<i>Parameter</i>	<i>Value</i>
Pixel	Rise/fall time	$\approx 120 \text{ ps}$
	Matched insertion loss	$\approx 45.5 \text{ } \mu\text{dB}$
	FWHM	$\approx 215 \text{ ps}$
	Pixel dimensions	$\approx 50 \text{ } \mu\text{m} \times 50 \text{ } \mu\text{m}$
	Quiescent current	$\approx 25 \text{ } \mu\text{A}$
	Dynamic power	$49 \text{ } \mu\text{W}$
	Pixel driver resistance	$\approx 3 \text{ G}\Omega$
	Pixel driver capacitance	$\leq 25 \text{ fF}$
Transmission line	Substrate ϵ_r	≥ 11.7
	Overall loss	$\leq 5.5 \text{ dB}$
	Δl_{PtoT_min}	$\approx 36 \text{ mm}$
Readout ASIC	Sampling speed	20 GS/s
	Number of channels	126
	Analog bandwidth	3 GHz
	Input referred noise	$\leq 0.5 \text{ mV}_{RMS}$
	Number of bits	12
	Power consumption per channel	$\approx 24.6 \text{ mW}$
	Data rate per chip	between 35 Mbits/s to 250 Mbits/s
	Added jitter	$\approx 25 \text{ fs}$

Ultimately, the overall timing performance depends on the actual resolution of the digitized waveform and the extraction algorithm used. In chapter four, a femtoresolution study in terms of the waveform sampling technique is presented exploring the various effects and mechanisms that contribute to the overall timing resolution of the system.

Considering the radiation hard environment of the detector, the radiation damage mitigation techniques [47] will be used in the detector version of the RFPix. For the first prototype, the specific techniques have been omitted in favor of focusing the design efforts to reach the desired performance.

Other notable challenges and aspects of the detector design that will not be addressed further in this work are:

- Active pixel array design and integration;

- Transmission line integration details into the substrate;
- Mechanical aspects:
 - Integration details such as the alignment of the subparts,
 - Mechanical aspects of cooling. The total power consumption of the sensor ladder is estimated at $\approx 25.8\text{ W}$ or 2.58 W/cm^2 :
 - * 6.3 W from two readout ASICs,
 - * $\approx 19.5\text{ W}$ from the pixel array.

CHAPTER 3

PSEC4 ANALYSIS RESULTS

An existing ASIC based on the waveform sampling technique that also achieves high sampling speeds and high timing resolutions is the PSEC4 [43] waveform digitizer. The PSEC4 is designed and fabricated in the IBM 130 *nm* node. The main specifications of the PSEC4 are listed in Table 3.1.

Table 3.1: PSEC4 characteristics

<i>Parameter</i>	<i>Value</i>
Sampling speed	$\leq 15 \text{ GS/s}$
Analog bandwidth	$\approx 1.5 \text{ GHz}$
Number of channels	6
Power consumption per channel	$\approx 10 \text{ mW}$

The sampling array of each PSEC4 channel is composed of 256 parallel switched capacitor cells (SCC). These cells continuously sample the input signal at the nominal sampling speed of $\approx 10 \text{ GS/s}$. The input signal is detected by a comparator, which issues a trigger signal. Following the trigger signal, a digitization cycle begins, where all of the sampling cells are digitized in parallel using a single slope analog-to-digital converter (ADC) scheme [48]. Alternatively, the trigger signal can initialize a transfer cycle of the sampled voltages into an analog storage array, which serves as an analog buffer, thus providing a very high information density. In both cases, the transfer rates are considerably lower than the sampling speed. In general, the transfer speed reduction factor equals to the sampling speed divided by the number of cells. However, that only holds true if all of the sampling cells begin the transfer cycle at the same time. Different transfer schemes have different reduction factors depending on the requirements. In other words, the sampling block is the only part of the ASIC that needs to run at the nominal sampling speed. In the RFpix case, the target sampling speed is 20 GS/s , hence requiring special consideration.

In the next sections, analysis results of the PSEC4 sampling cell are presented. The underlying mechanisms that limit the performance are identified and quantified. Contingent upon these results, the RFpix SCA architecture is presented along with simulation results showing the RFpix SCC performance.

3.1 Basic Design Scheme and Layout

The PSEC4 SCA has a single-ended configuration of the input transmission line. Consequently, the individual sampling cells also follow a single-ended configuration.

Each cell is composed of an input CMOS switch followed by a sampling capacitor and a comparator. The comparator serves as the input stage of the slope ADC. Figure 3.1 shows a simplified block schematic of a single PSEC4 SCA sampling cell.

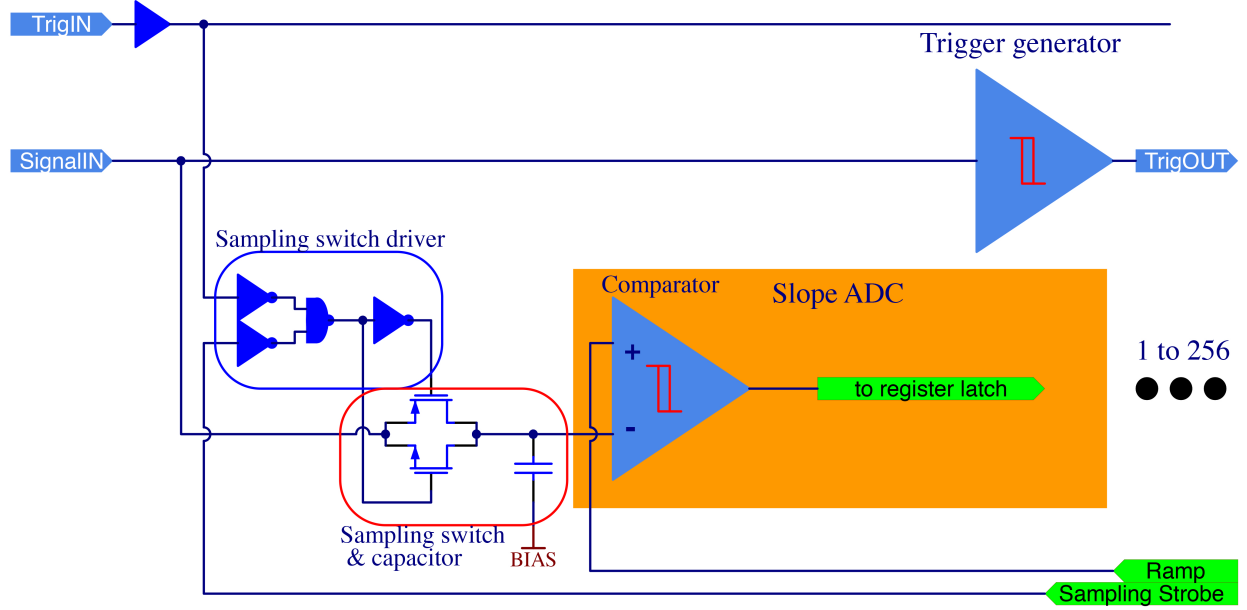


Figure 3.1: Block diagram of a single PSEC4 SCA sampling cell.

The layout of a PSEC4 input channel is shown in Figure 3.2a while the layout of the input sampling cell is shown in Figure 3.2b

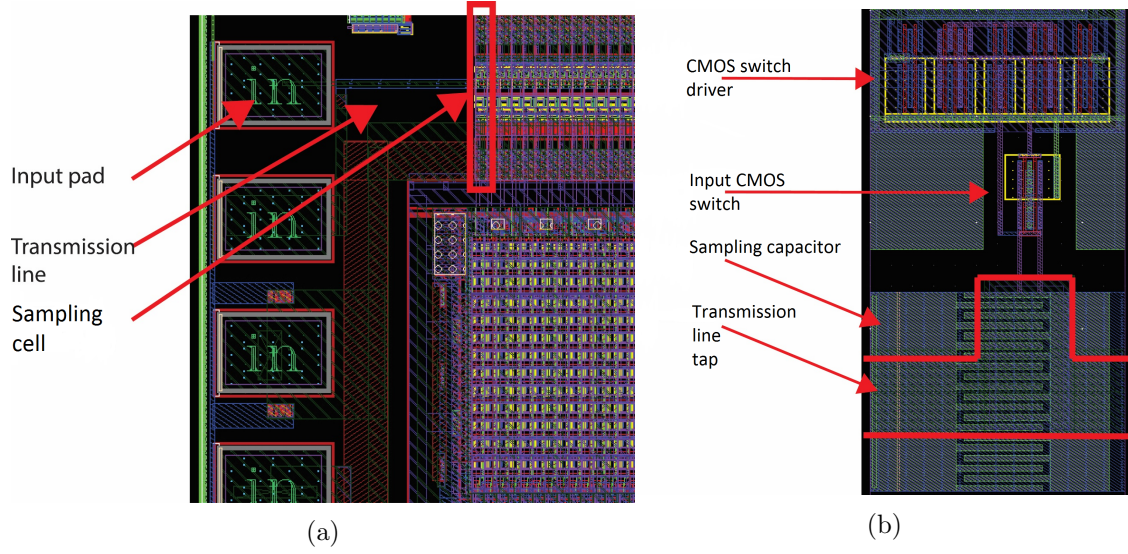


Figure 3.2: (a) PSEC4 input channel layout with input pad, transmission line, and sampling cell. (b) Layout of a single PSEC4 sampling cell showing the switch driver, input CMOS switch, sampling capacitor, and the transmission line tap.

3.2 DC Analysis and Characterization of Basic Components

The basic components making up the cell are: a CMOS switch, a sampling capacitor, and the comparator. It has to be noted that all of the presented simulation results have been acquired through the Cadence Virtuoso design environment. All of the simulation models include layout parasitics.

The input CMOS switch is composed of two complementary transistors, a PMOS and an NMOS. The equivalent circuit can be simplified to a first order CRC circuit as shown in Figure 3.3.

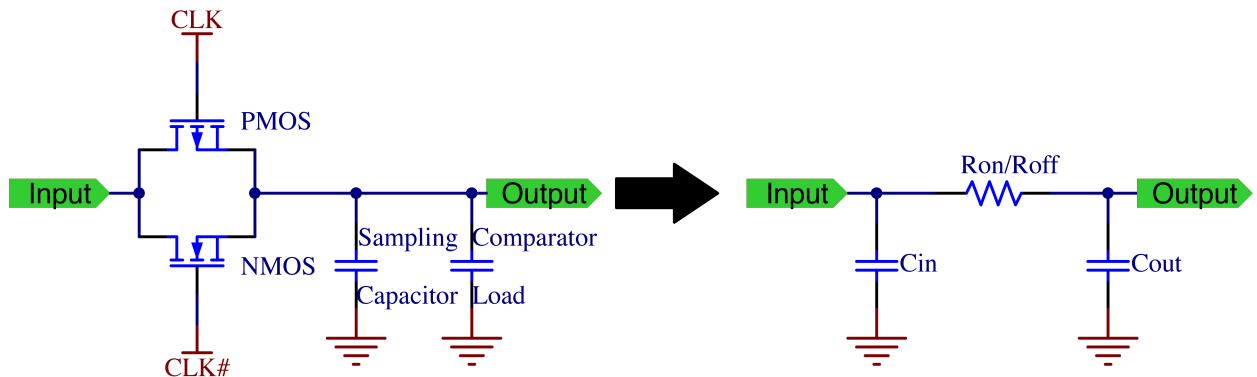


Figure 3.3: Single PSEC4 sampling cell with equivalent circuit.

3.2.1 Switch Resistance

The simulated Track mode (ON-state) and Hold mode (OFF-state) resistances are shown in Figure 3.4. The peak resistance in track mode is approximately $2.4 \text{ k}\Omega$ at $V_{DC} = 665 \text{ mV}$. The standard deviation over the dynamic range is 614.8Ω .

The hold mode resistance is in $G\Omega$, thus presenting good isolation. Considering the fact that one of the transistors is always open close to the rails, along with the signal requirements postulated in chapter two, and the eventual saturation limits of the buffers close to the rails, it has been decided that the signal dynamic range in the RFpix will span from 100 mV to 1100 mV .

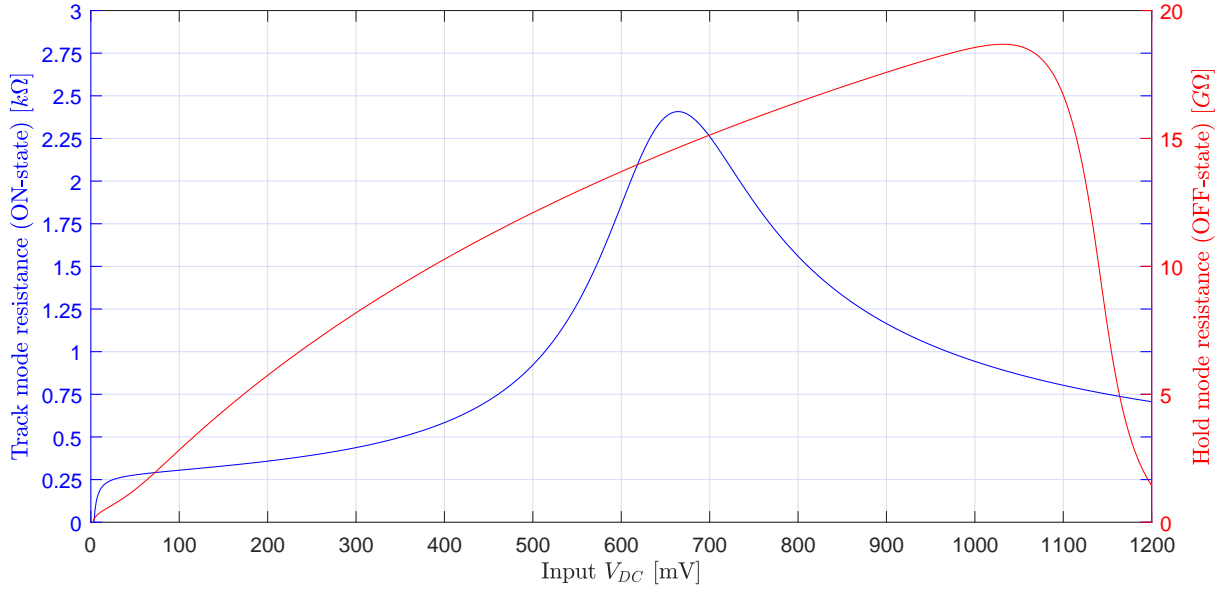


Figure 3.4: Track mode (ON-state) and Hold mode (OFF-state) resistances as functions of input V_{DC} .

3.2.2 Sampling Cell Capacitances

The input (C_{IN}) and output (C_{OUT}) capacitances of the circuit shown in Figure 3.3 have been determined by simulating the input impedance (Z_{11}) of the circuit and extracting the values of the individual terms. In the Laplace domain, Z_{11} can be written as shown in Equation 3.1.

$$Z_{11} = \frac{1 + sC_{OUT}R_{ON/OFF}}{s^2C_{IN}C_{OUT} + s(C_{IN} + C_{OUT})} \quad (3.1)$$

By substituting s with $j\omega$ and rearranging the terms so that the real and imaginary parts are separated ($Z_{11} = Re + jXe$), the reactance can be written as shown in Equation 3.2.

$$X_{e11} = \frac{\omega(C_{IN} + C_{OUT}) + \omega^3(C_{IN}C_{OUT}^2R_{ON/OFF}^2)}{\omega^2(C_{IN} + C_{OUT})^2 + \omega^4(C_{IN}^2C_{OUT}^2R_{ON/OFF}^2)} \quad (3.2)$$

The negative polarity of the reactance tells us that it is capacitive. This capacitance is determined by the relation shown in Equation 3.3.

$$C = \frac{1}{\omega \text{Im}(Z_{11})} \quad (3.3)$$

Considering the limiting cases when the frequency goes to either zero or infinity as shown in Equations 3.4 and 3.5 respectively, all of the capacitances can be determined as shown in Equations 3.4 and 3.5.

$$\lim_{\omega \rightarrow 0} X_{e11} = \frac{\omega(C_{IN} + C_{OUT})}{\omega^2(C_{IN} + C_{OUT})^2} = \frac{1}{\omega(C_{IN} + C_{OUT})} \quad (3.4)$$

$$\lim_{\omega \rightarrow \infty} X_{e11} = \frac{\omega^3(C_{IN}C_{OUT}^2R_{ON/OFF}^2)}{\omega^4(C_{IN}^2C_{OUT}^2R_{ON/OFF}^2)} = \frac{1}{\omega C_{IN}} \quad (3.5)$$

The right side capacitance of the model (Figure 3.3) denoted as C_{OUT} is composed of the transistor output capacitance, the sampling capacitor, and the comparator load capacitance. The left side capacitance of the model denotes as C_{IN} is the input transistor capacitance. Figure 3.5 shows the extrapolated capacitances simulated under different load conditions as functions of frequency.

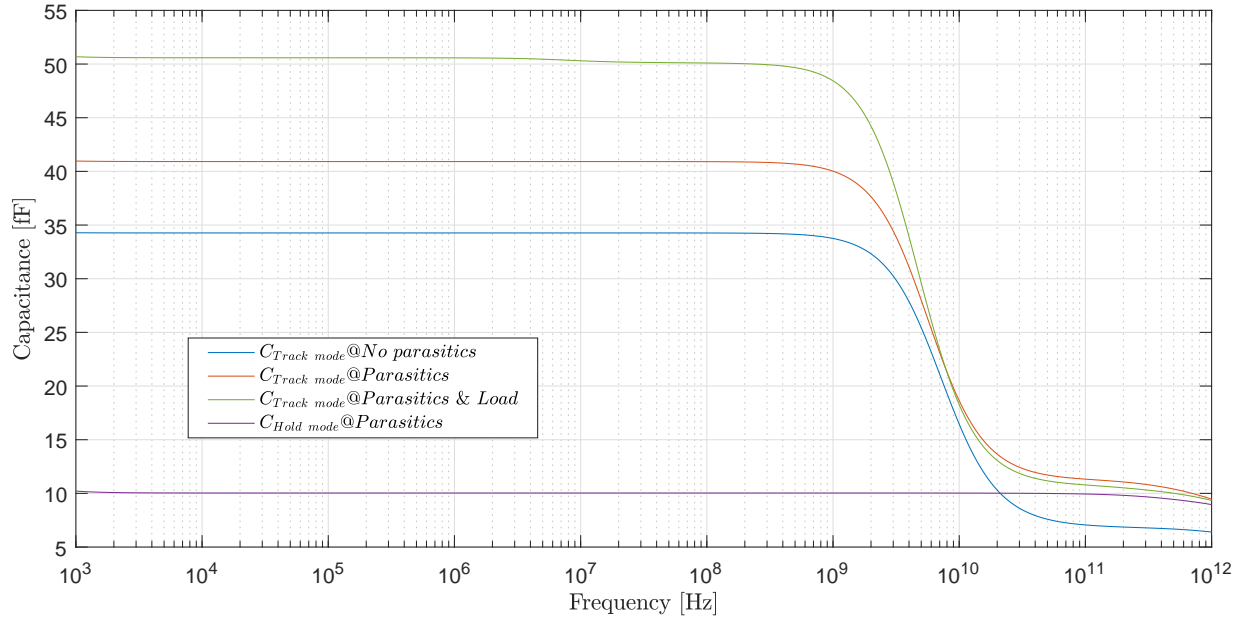


Figure 3.5: Extrapolated capacitance as a function of frequency.

Using the values shown in 3.5, all of the sampling cell capacitances have been accounted for and are summarized in Table 3.2.

Table 3.2: Extrapolated PSEC4 sampling cell capacitances.

<i>Capacitance</i>	<i>Value [fs]</i>
$C_{IN\ TRAN.}$	10.0
$C_{OUT\ TRAN.}$	10.0
$C_{SAMPLING}$	20.3
C_{LOAD}	11.0

3.3 Frequency Response of the PSEC4 Sampling Cell

3.3.1 Tracking Bandwidth

The tracking bandwidth of the sampling cell is the bandwidth needed by the sampling cell in order to be able to follow the input signal changes in time. This bandwidth is completely defined by the time constant (RC) formed by the switch track mode resistance and the inner sampling cell capacitance. In the light of the switch track mode resistance variance, a similar response is expected of the tracking bandwidth. Figure 3.6 shows the bandwidth and the track mode switch resistance as functions of the input DC voltage (V_{DC}). As expected, the lowest bandwidth occurs at $V_{DC} \approx 650\ mV$, where the bandwidth is approximately $1.68\ GHz$.

3.3.2 Group Delay

Group delay (GD) gives us insight into the rate of change of the signal phase with respect to frequency. It is thus defined as: $GD = -d\phi(\omega)/d\omega$. In other words, it is a measure of how much the individual frequency components are delayed in time in respect to the input. Hence, it gives us the first insight into signal distortion. Due to the large variance in bandwidth, the GD was also simulated in terms of input V_{DC} . Figure 3.7 shows the GD as a function of the input V_{DC} and frequency. Close to the low bandwidth point the GD reaches as high as $90\ ps$.

3.4 Large Signal Analysis

With such a large variance of the tracking bandwidth and group delay, distortion effects are expected. It has to be noted that all of the following large signal simulation results take into account 30 harmonics.

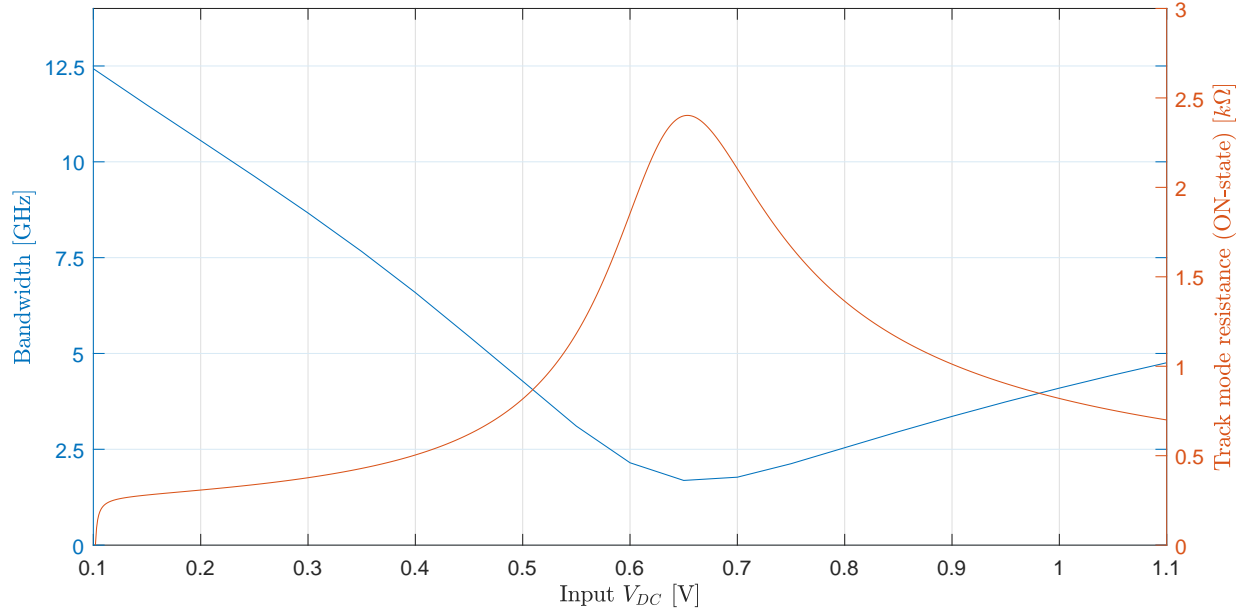


Figure 3.6: Bandwidth and track mode switch resistance as functions of Input V_{DC} .

Figures 3.8a and 3.8b show the signal attenuation of the sampling cell as a function of signal voltage, frequency, and input V_{DC} . In both figures, the attenuation of the signal has a dependence on the input V_{DC} , where the highest attenuation occurs around $V_{DC} = 650 \text{ mV}$. Furthermore, a dependency on the frequency and the signal amplitude is also observed. Further, Figure 3.8c shows the total harmonic distortion as a function of signal amplitude and input V_{DC} . There are many definitions for the THD. Equation 3.6 shows the one used through this work.

$$THD = \frac{\sqrt{\sum_{n=2}^{30} V_n^2}}{V_1} [\%] \quad (3.6)$$

The signal attenuation observed at small signal amplitudes can be explained by taking a closer look at the attenuation and distortion curves for a particular input V_{DC} . In our case, we chose a $V_{DC} = 600 \text{ mV}$. Figure 3.9 shows the total harmonic distortion (THD) and attenuation as functions of the signal amplitude at $V_{DC} = 600 \text{ mV}$ and $f_{SIG} = 1 \text{ GHz}$. Three regions of operation can be clearly identified:

1. **Low THD and high attenuation:** At small signal amplitudes, the signal distortion is low, because the variance of track mode switch resistance is small. Conversely, the signal compression is high, which correlates with the minimum bandwidth shown in Figure 3.6 for $V_{DC} \approx 650 \text{ mV}$. That is, the signal attenuation is a consequence of the limited tracking bandwidth.
2. **Medium THD and low attenuation:** Due to varying resistance of the switch over dy-

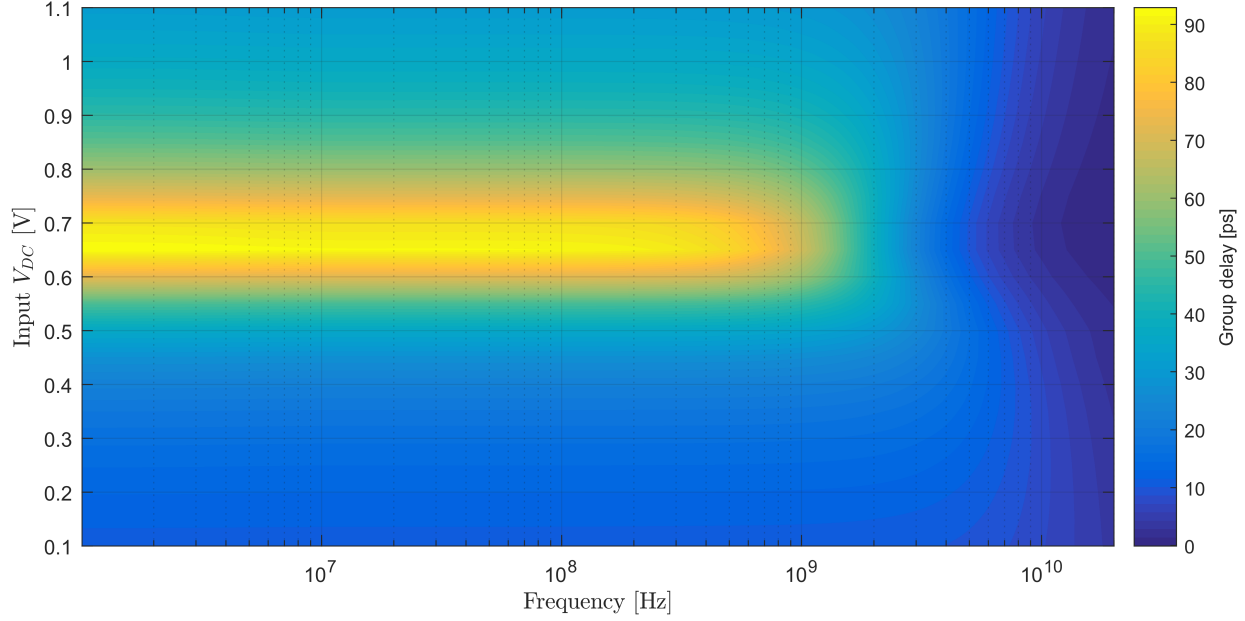


Figure 3.7: Sampling cell group delay as a function of input V_{DC} and frequency.

dynamic range, the output signal will be delayed and attenuated differently depending on the instantaneous voltage on the input. This non-linear response gives rise to harmonics, increasing the distortion. However, the harmonics are less delayed at lower switch resistance values. Consequently, most of these harmonics interfere constructively with the fundamental tone, which increases the amplitude further away from the $V_{DC} = 650 \text{ mV}$ point. However, the THD increases because of these same harmonics.

3. **High THD and high attenuation:** At high signal amplitudes, both the distortion and the attenuation are highest due to the large track mode switch resistance variance and the compression resulting from the proximity of the supply rails.

3.5 Noise and ENOB

3.5.1 Noise

Noise is a fundamental property of every electronic circuit. There are various types of noise models depending on their origins and energy spectrum shape. In our case, the two dominant noise types are the flicker noise, also called $1/f$ noise, and the thermal noise [49].

Flicker noise dominates at lower frequencies and can be calculated as:

$$\overline{v_n^2} = \frac{K}{C_{OX}WL} \cdot \frac{1}{f} \quad (3.7)$$

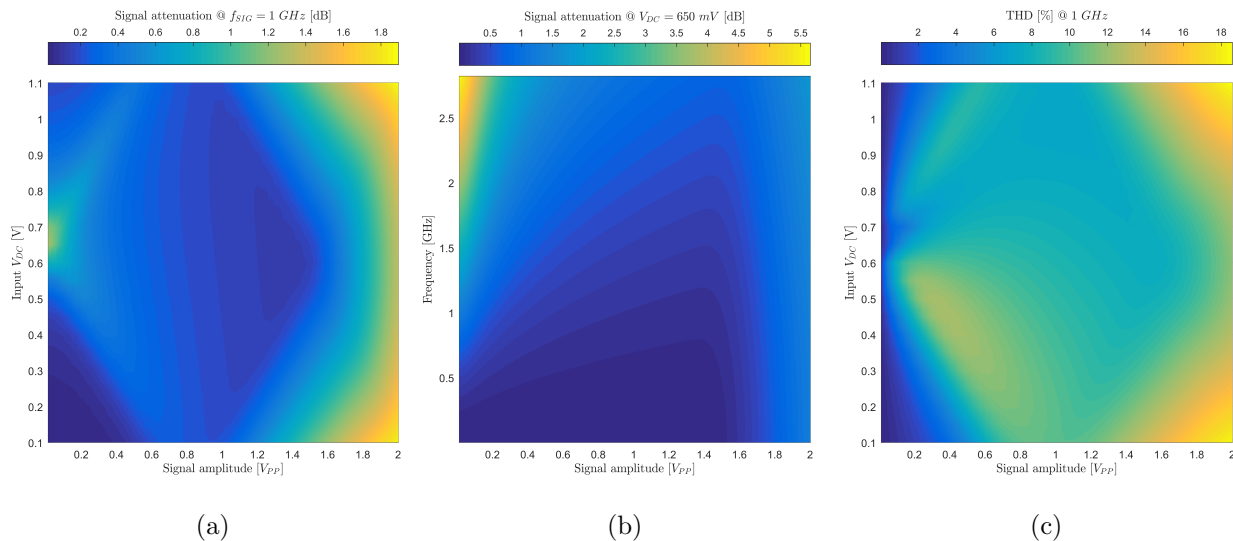


Figure 3.8: (a) PSEC4 sampling cell signal attenuation as a function of input V_{DC} and signal amplitude at a signal frequency of 1 GHz . (b) PSEC4 sampling cell signal attenuation as a function of frequency and signal amplitude at an input V_{DC} of 650 mV . (c) Total harmonic distortion as a function of input V_{DC} and signal amplitude.

However, due to our frequency range, the impact of this noise type is negligible and it will be neglected in our analysis.

JohnsonNyquist noise (thermal noise, Johnson noise, or Nyquist noise) is the electronic noise generated by the thermal agitation of the charge carriers (usually the electrons) inside an electrical conductor at equilibrium, which happens regardless of any applied voltage. The generic, statistical physical derivation of this noise is called the fluctuation-dissipation theorem, where generalized impedance or generalized susceptibility is used to characterize the medium.

Thermal noise of an ideal resistor is approximately white, meaning that the power spectral density is nearly constant throughout the frequency spectrum. When limited to a finite bandwidth, thermal noise has a nearly Gaussian amplitude distribution [50]. Figure 3.10 shows the output referred noise (ORN) of the sampling cell as a function of frequency and input V_{DC} .

The noise is highest when the switch resistance is highest, thus suggesting that the thermal noise of the switch track mode resistance is dominating the noise spectrum. At higher frequencies, the noise starts to drop due to the bandwidth limit.

3.5.2 ENOB

The effective number of bits (ENOB) is a figure of merit used to characterize the effective vertical resolution of a system. The ENOB can be obtained by using Equation 3.8 [51],

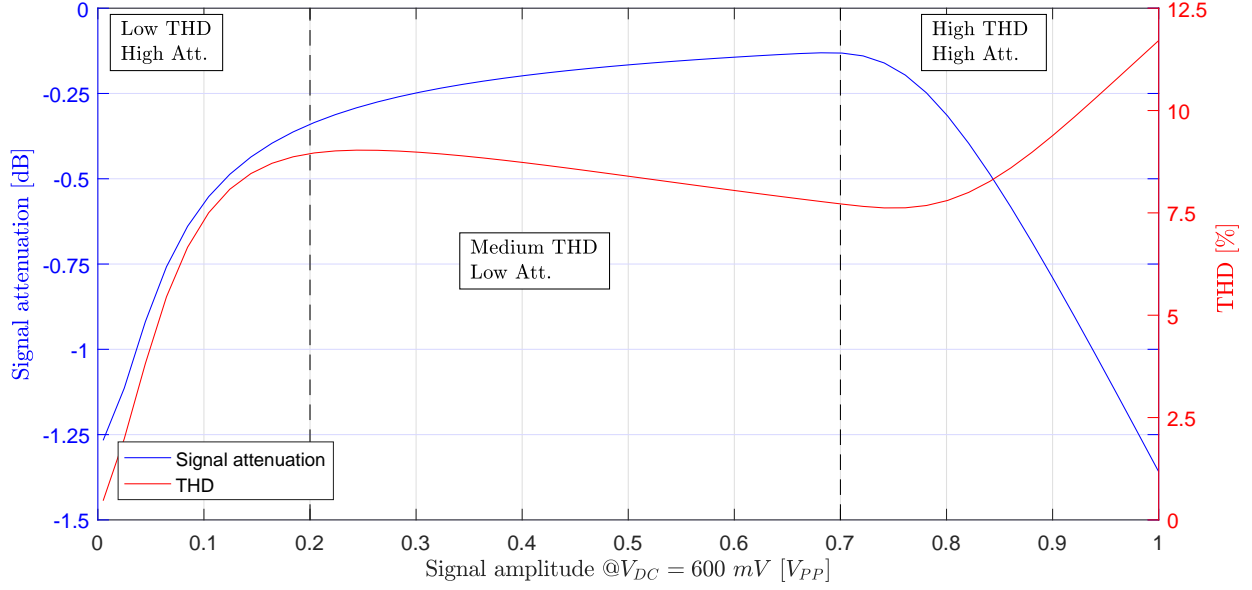


Figure 3.9: Distortion and signal attenuation as functions of signal amplitude at input $V_{DC} = 600$ V.

$$ENOB = \frac{SINAD - 1.76 + 20\log_{10}\left(\frac{Fullscale\ voltage}{Signal\ voltage}\right)}{6.02} \quad (3.8)$$

SINAD is the signal-to-noise-and-distortion ratio which takes into account the noise of the system and distortion. The ratio is given by Equation 3.9,

$$SINAD = \frac{P_{SIGNAL} + P_{NOISE} + P_{DISTORTION}}{P_{NOISE} + P_{DISTORTION}} \quad (3.9)$$

where the terms are the respective powers of the signal, noise, and distortion in watts. A more practical expression is shown in 3.10,

$$SINAD = -10\log_{10}\left[10^{-\frac{SNR[dB]}{10}} - THD[\%]\right] \quad (3.10)$$

where SNR is the signal to noise ratio. The third term in the numerator of Equation 3.8 corrects the equation for measurements that are not done at full-scale. That is, if the input amplitude is not at full-scale, the SNR decreases and consequently the SINAD. In our case, the full-scale is 1 V_{PP} . Figures 3.11a and 3.11b show the ENOB of the PSEC4 sampling cell as a function of input V_{DC} , signal amplitude, and frequency.

The overall ENOB in this case is almost completely dominated by distortion. However, calibration procedures can counter this effect to some extent, though not completely. In order to achieve the required performance, the RFpix input switch is being designed towards low variance of the track mode resistance.

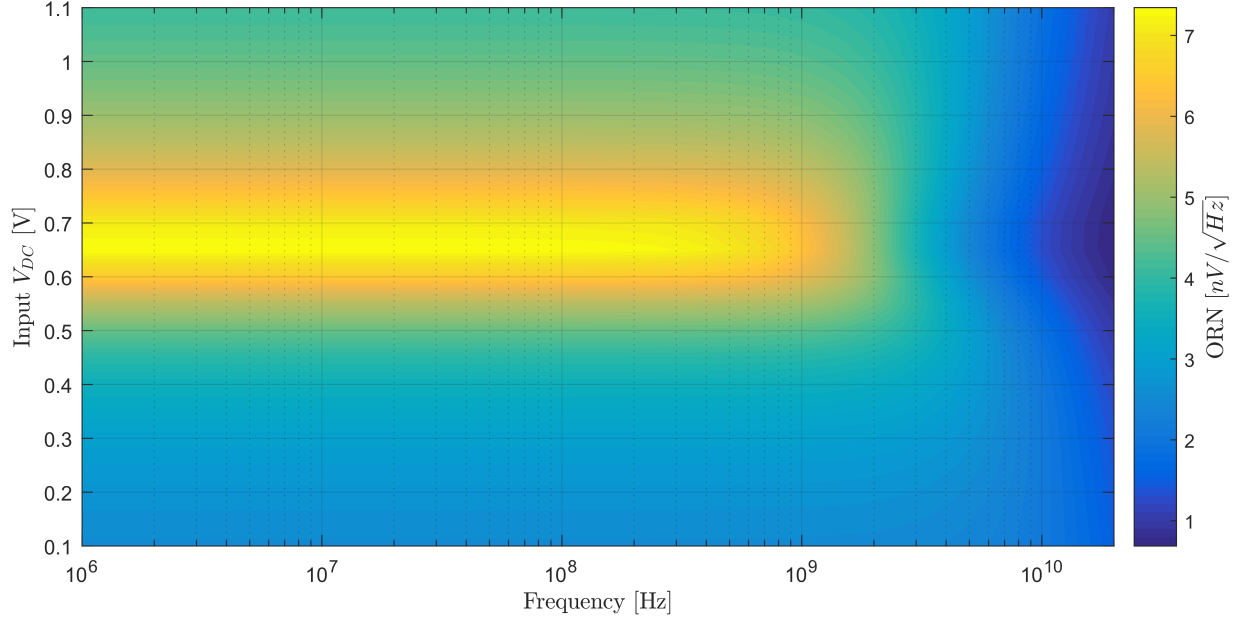


Figure 3.10: Output referred noise as a function of input V_{DC} and frequency.

3.6 Transient Response

So far we explored steady state responses. In this section, simulation results are presented showing the transient response of the sampling cell. Here we determine the achievable speed of the sampling cell, show the artifacts of the transition (forward transient and kickback) as well as acquire insight into the pedestal error mechanisms, mainly charge injection and clock feedthrough.

When in track mode a conduction channel exist at the oxide-silicon interface of switch transistors. When the switch transitions into hold mode, the charge in the channel exits through the source and the drain terminals, the phenomenon is thus called “channel charge injection”. The amount of charge in the channel depends on the input voltage and can be calculated by using Equation 3.11,

$$Q_{CH} = WLC_{OX}(V_{DD} - V_{IN} - V_{TH}) \quad (3.11)$$

where Q_{CH} is the charge in the channel, W and L are the transistor width and length respectively, C_{OX} is the oxide capacitance, V_{DD} is the supply voltage, V_{IN} is the voltage at the input of the switch, and V_{TH} is the transistor threshold voltage. The charge injected to the input of the switch is in general absorbed by the line driver. The charge that is injected to the output of the switch is deposited to the sampling capacitor. The voltage on the sampling capacitor is not driven in hold mode and it is directly defined by capacitance (C) and charge (Q) as $V = Q/C$. So extra charge from the channel induces a voltage error.

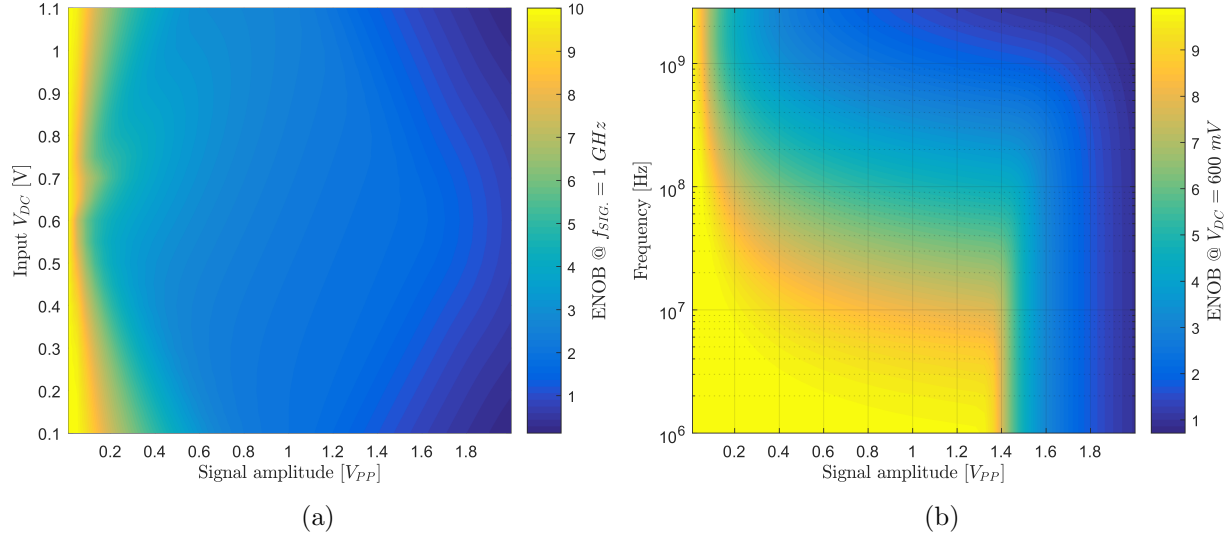


Figure 3.11: (a) ENOB of the PSEC4 sampling cell as a function of input V_{DC} and signal amplitude. (b) ENOB of the PSEC4 sampling cell as a function of frequency and signal amplitude.

In summary, charge injection contributes three types of errors in MOS sampling circuits [52]:

1. **Gain error:** Charge injected due to channel capacitance:

$$Q_{CH} = WLC_{OX}(V_{DD} - V_{IN} - V_{TH}) \quad (3.12)$$

This error depends on the input voltage and it manifests as gain error.

2. **DC offset:** Charge injection due to gate overlap capacitance:

$$Q_{CL} = -WL_{OV}C_{OX}(V_{DD} - V_{SS}) \quad (3.13)$$

This error is independent of the input voltage, hence manifesting as a static voltage offset.

3. **Non-linearity:** Charge injection due to body effect (threshold voltage variation depended on frequency). This error is non linearly related to the input voltage which causes distortion.

In many applications, the first two can be tolerated or corrected, whereas the last one cannot be [52].

In addition to channel charge injection, a MOS switch couples the clock transitions to the sampling capacitor through its gate-drain or gate-source overlap capacitance. The error is independent of the input level, manifesting itself as a constant offset in the input/output characteristic. As with charge injection, clock feedthrough leads to a trade-off between speed and precision [49].

The voltage offset error is also called the pedestal error.

Figures 3.12a, 3.12b, and 3.12c show the transient response of the PSEC4 sampling cell for three input V_{DC} voltages. Depending on this voltages the pedestal error voltage changes magnitude and polarity. Thus, it is a voltage dependent disturbance.

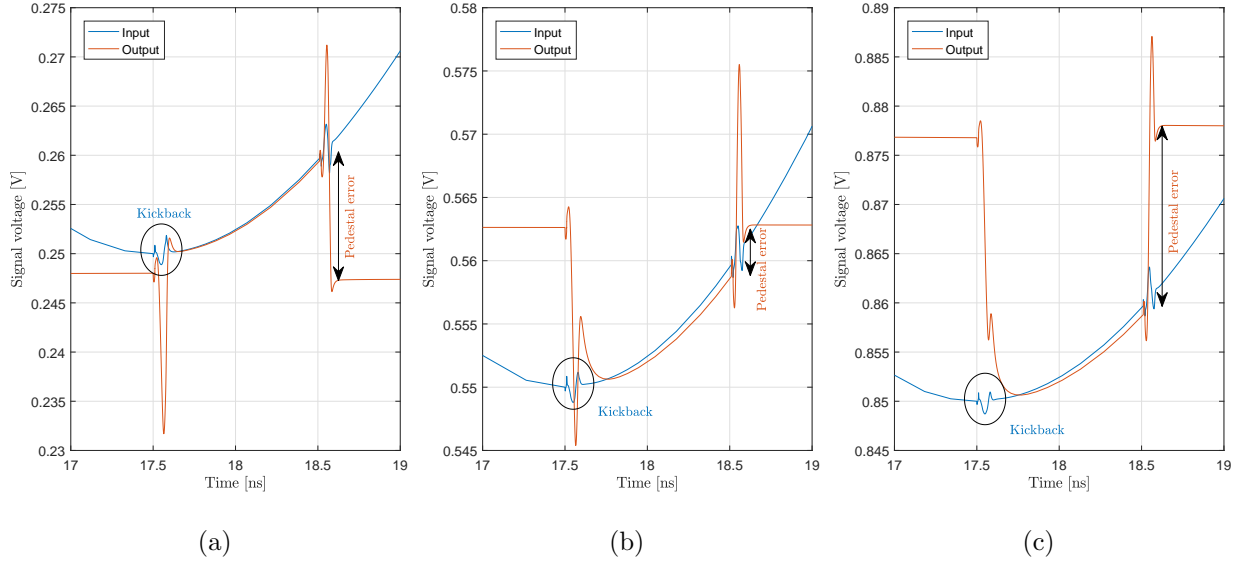


Figure 3.12: PSEC4 sampling cell transient response with 1 ns time window at three different input V_{DC} voltages: (a) $V_{DC} = 300 \text{ mV}$, (b) $V_{DC} = 600 \text{ mV}$, and (c) $V_{DC} = 900 \text{ mV}$.

Acquisition time is the time needed to switch from hold mode into track mode and for the sampling cell voltage to start following the input signal within a certain error. Settling time is the time needed from the onset of the hold clock transition to the eventual settling of the sampling cell voltage to a DC value. Table 3.3 shows the PSEC4 acquisition and settling times at three different input V_{DC} voltages.

Table 3.3: PSEC4 acquisition and settling times at three different input V_{DC} voltages.

<i>Input V_{DC}</i>	<i>Acquisition time</i>	<i>Settling time</i>
300 mV	0.14 ns	0.11 ns
600 mV	0.68 ns	0.11 ns
900 mV	0.52 ns	0.11 ns

The acquisition time is mostly dependent on the bandwidth and hence it is longer close to the $V_{DC} = 600 \text{ mV}$ where the bandwidth is lowest. Settling time depends on how quickly the charge settles on the sampling capacitor, hence it is independent of the bandwidth and it does not present any variance over the dynamic range.

3.7 Summary of PSEC4 Sampling Cell Analysis

The analysis of the PSEC4 sampling cell lead us to significant conclusions affecting the RFPix design. This conclusions are:

1. A CMOS switch has large variance of the track mode resistance, which causes a large variance in the tracking bandwidth. Thus, a sampling switch with considerably lower track mode resistance variance has been designed. The switch is presented in chapter five.
2. The distortion resulting from the large bandwidth variance is the dominant factor in limiting the ENOB of the system. From chapter two, we know that this number needs to be in the range of approximately 9.7 effective bits to achieve the necessary timing resolution. In chapter four we will discuss this issue in more detail.
3. The pedestal error is a voltage dependent disturbance of the sampled signal which degrades the signal considerably. Calibrating for the entire dynamic range is time consuming. Therefore, a more practical solution has been devised and presented in chapter five.

CHAPTER 4

FEMTOSECOND RESOLUTION TIMING IN MULTI-GS/S WAVEFORM DIGITIZING ASICS

Solutions that can achieve sub-picosecond timing resolutions are mostly limited to the applications in the optical domain [53, 54]. One class of exceptions are RF continuous wave (CW) synchronization systems [55]. These systems use RF phase detectors with the precision and stability of a few fs . Such phase detectors are either integrated circuits (ICs) [56] or they are composed of discrete components, where the principle of operation is based on the saturated mixer technique [31]. Their drawback is low bandwidth of operation, which inhibits the detection of rapid events.

The RFpix principle of operation as discussed in chapter two is based on the waveform sampling technique. This technique does not just give the best timing resolution performance for the given power consumption, but it also allows for fast measurements.

The design of the RFpix ASIC requires very good understanding of the effects that impact the timing resolution. This chapter presents simulation results that clearly identify and quantify the sources of error and the underlying coupling mechanisms. In addition, a synthetic waveform generator developed solely for this purpose, is presented and validated through measurement results.

4.1 Waveform Sampling Technique

As explained in section 2.3.1, the timing extraction is based on linear interpolation between two waveform samples, which assumes a linear interpolation for Equation 2.8. Using the ASIC design specifications given in Table 2.3 as inputs to Equation 2.8, the theoretically achievable timing resolution is estimated at approximately $37 fs$. Direct testing of this prediction presents a challenging task considering that, to the authors' knowledge, such an IC is not yet commercially available. In light of this, the SWG has been developed in the MATLAB[®] software, with the intention to provide sampled signal emulation as realistically as possible. Furthermore, the SWG has been designed specifically to enable the separation and study of the various sources of error, along with their coupling mechanisms in order to understand, quantify, and evaluate the individual contributions to the overall timing resolution.

4.2 Synthetic Waveform Generator

The SWG currently supports three signal shapes: a Gaussian pulse, a square pulse, and a sine wave. Other SWG features include:

- Synchronous vs asynchronous modes of operation with the options to choose phase distribution the phase offset;

- Amplitude noise injection;
- Quantization with adjustable number of bits (N_{BITS});
- Jitter injection;
- Adjustable sampling frequency;
- Adjustable signal frequency and bandwidth;
- Adjustable time window;
- Adjustable full scale dynamic range.

The generator output provides four vectors:

- Sampled amplitude vector (*Samples*);
- Sampler time base vector with jitter (t_{SAMP});
- Sampler time base vector without jitter (t_{SAMP} *w/o jitter*);
- Signal time base vector (t_{SIGNAL}).

These four vectors represent the signal waveform analogous to the output of an actual physical waveform digitizer.

Figure 4.1 shows the block diagram of the SWG.

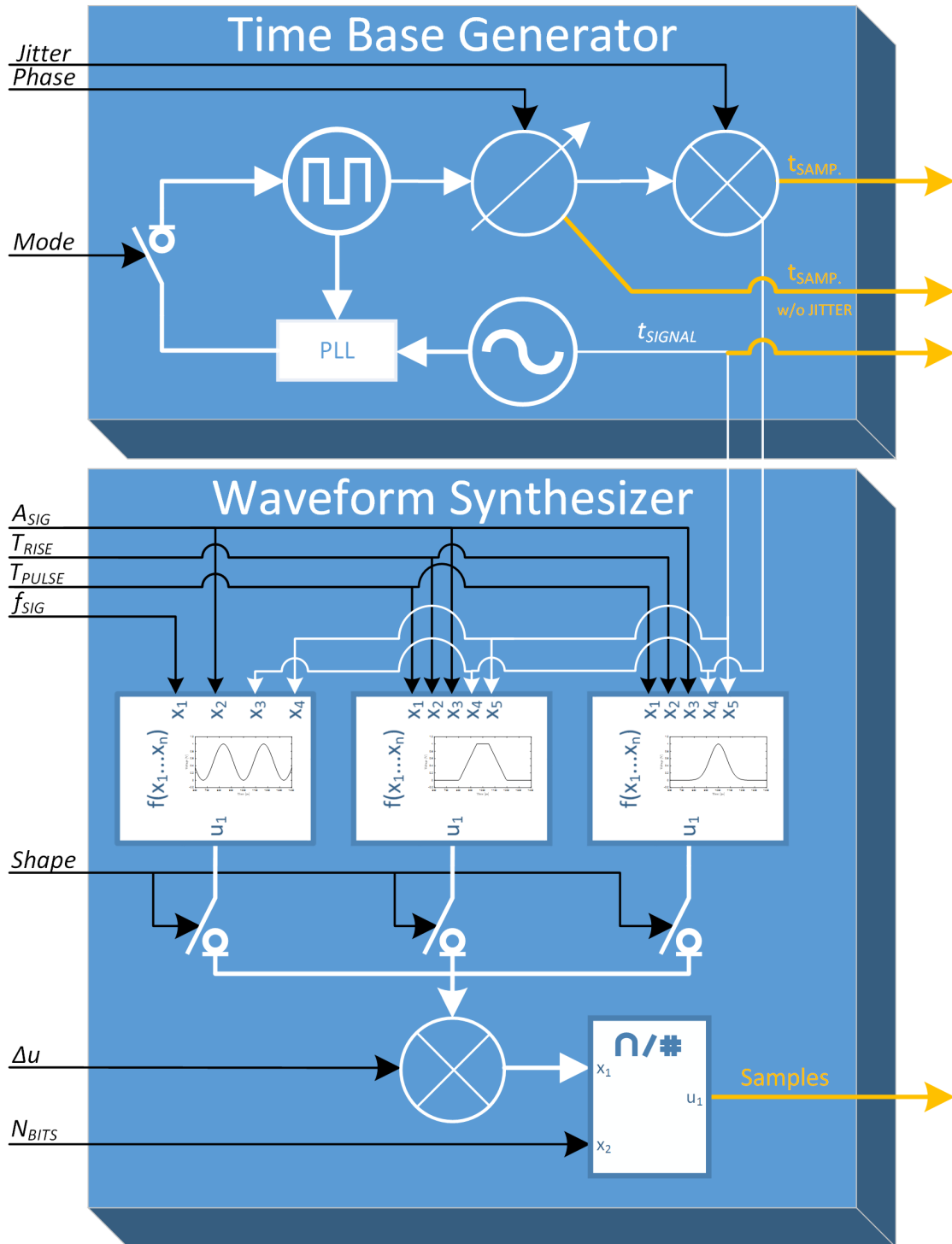


Figure 4.1: Functional block diagram of the SWG code with the time base generator and the waveform synthesizer substructures. Input variables are marked in black on the left side, while the output vectors are marked in yellow on the right side.

4.3 Case Studies and Analysis

Various algorithms can be used to extract the time of arrival from a sampled waveform. However, to test the predictions of Equation 2.8, a linear fit needs to be used. Furthermore, a linear fit only requires two samples on the leading edge in order to successfully interpolate the transition time. This minimizes the data rate which results in processing power reduction. Equation 4.1 shows the calculation used for the linear fit between two adjacent samples,

$$t_{INT.} = \frac{(A_{TH.} - A_i) \times (t_{i+1} - t_i)}{A_{i+1} - A_i} + t_i \quad (4.1)$$

where $t_{INT.}$ is the interpolated transition time through the chosen threshold voltage $A_{TH.}$. Unless otherwise specified, the threshold is set to half of the full signal amplitude ($A_{TH.} = U/2$). A_i , A_{i+1} are the amplitudes of the adjacent samples (S_i , S_{i+1}) over the chosen threshold and t_i , t_{i+1} are their corresponding time stamps. Figure 4.2 shows an example of a noisy Gaussian pulse waveform generated by the SWG with an overlay visualizing the linear fit. The parameters used to generate this waveform are the same as given in Table 2.3, with the exception of the noise. In order to make the effect more visible, the injected noise is 10 mV_{RMS} .

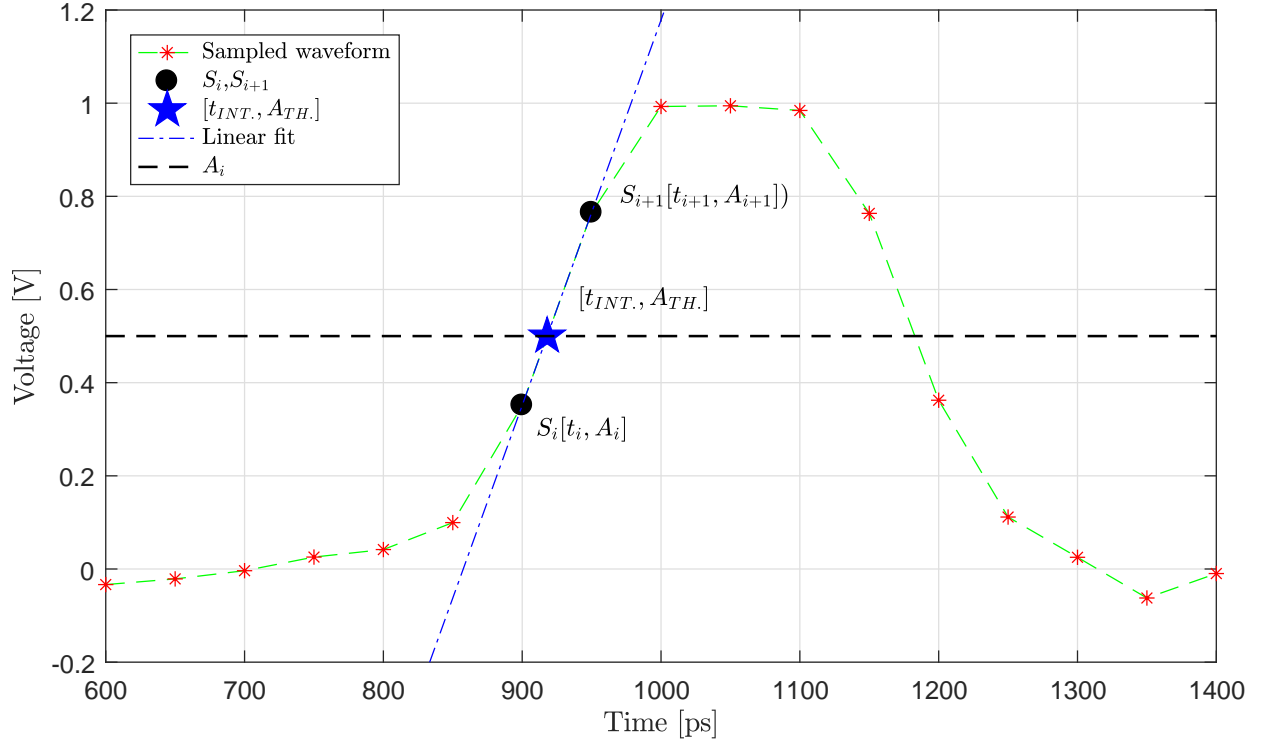


Figure 4.2: Example of a noisy Gaussian pulse waveform generated by the SWG with an overlay visualizing the linear extraction algorithm.

To comprehensively study the effects and mechanisms that can be simulated with the SWG, several case studies have been devised:

- Signal shape effect on the extraction algorithm;
- Synchronous vs asynchronous mode of operation;
- Amplitude noise vs quantization;
- Jitter injection;
- Scaling of the interpolated transition time resolution ($\sigma_{t_{INT.}}$) as a function of the $f_{SAMP.}$, BW , and Δu .

Unless otherwise specified, the parameter values used for signal generation are the ones given in Table 2.3.

4.4 Simulation Results

4.4.1 Synchronous vs Asynchronous Mode and Signal Shape Impact

In the synchronous mode, the time base of the signal is phase-locked to the time base of the sampler, while in the asynchronous mode it is not. That is, in the synchronous mode of operation, the sampling speed is an integer multiple of the signal rate. When unlocked, the phase difference between the two is completely random and thus modeled by a uniform distribution in the SWG.

The results summarized in Table 4.1 indicate that the asynchronous mode provides worse performance in all cases compared to the estimated value of $37 fs$. The random normal distribution of the phase difference coupled with the inherent error of the linear fit versus the non-linear waveform edges dominate the $\sigma_{t_{INT.}}$ in both, the sine and Gaussian, waveforms. In the case of the square pulse, the edge is linear and thus the random distribution of the phase difference does not significantly degrade the $\sigma_{t_{INT.}}$.

Table 4.1: $\sigma_{t_{INT.}}$ in the synchronous and asynchronous modes of operation for all three signal types.

<i>Mode</i>	<i>Signal Shape</i>		
	<i>Sine</i>	<i>Square</i>	<i>Gauss</i>
Synchronous	61 <i>fs</i>	50 <i>fs</i>	44 <i>fs</i>
Asynchronous	543 <i>fs</i>	61 <i>fs</i>	678 <i>fs</i>

Figure 4.3 shows an example of the positioning of the sampling points S_i , S_{i+1} over the threshold voltage for both, the synchronous (left) and asynchronous (right), modes of operation.

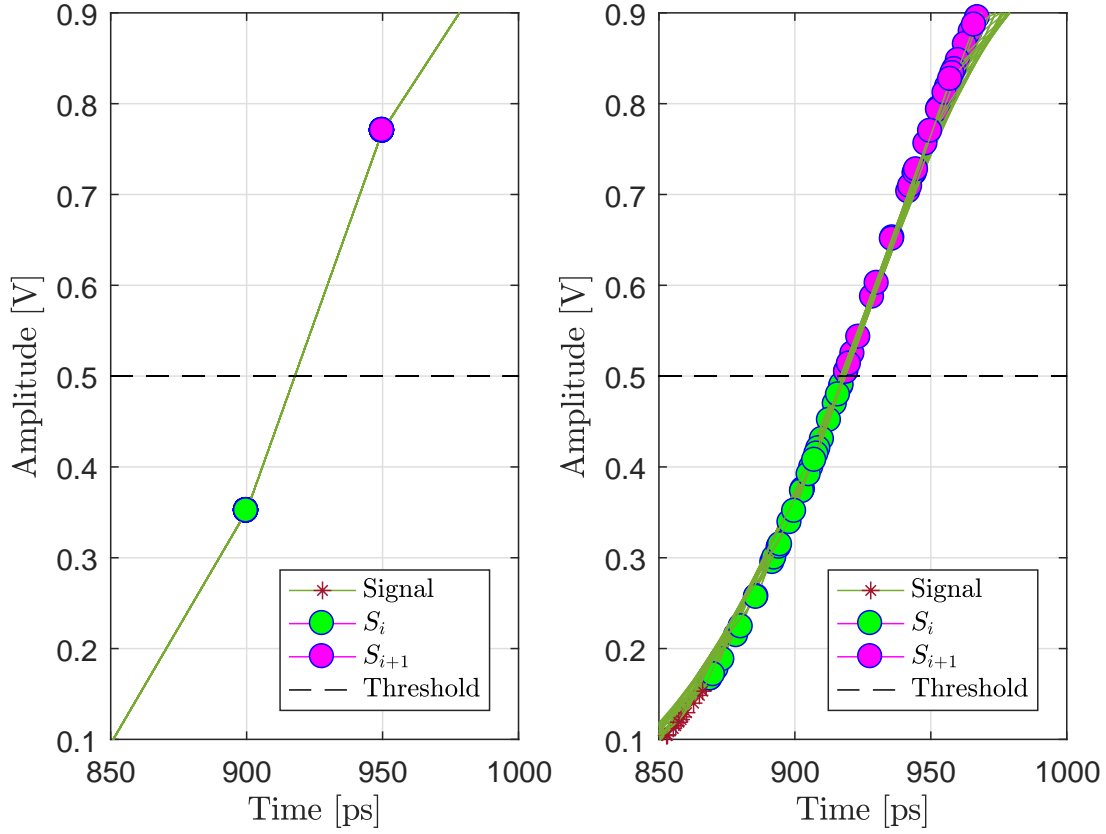


Figure 4.3: Synchronous (left) and asynchronous (right) modes of operation ($N_{acq.w.} = 30$).

Further investigation indicates that the degradation of the $\sigma_{t_{INT.}}$ is affected by the changing asymmetry of the sampling interval over the threshold voltage. Figure 4.4 shows a persistence plot of 500 fitted waveforms over the threshold voltage.

When the interval is perfectly symmetric, the probability of the fitted line crossing the threshold at the same time is the highest, thus improving the $\sigma_{t_{INT.}}$. In contrast, when the threshold is close to one of the sampling points, the timing ambiguity caused by the amplitude noise is the highest. Or, in other words, the amplitude modulation to phase modulation (AM-to-PM) conversion factor is the highest close to the sampling points. This effect is masked in the asynchronous mode by the random phase difference between the two time bases, consequently smearing the measurements. For the purpose of quantifying the asymmetry of the sampling interval over $S_{INT.}$, we define the SIF parameter. The mathematical definition of this figure of merit (FOM) is given in Equation 4.2.

$$SIF = 100 \cdot \frac{t_{i+1} + t_i - 2 \cdot t_{INT.}}{t_{i+1} - t_i}. \quad (4.2)$$

At the same time, it is worth defining the mean of the SIF (μ_{SIF}). That is, the mean of the

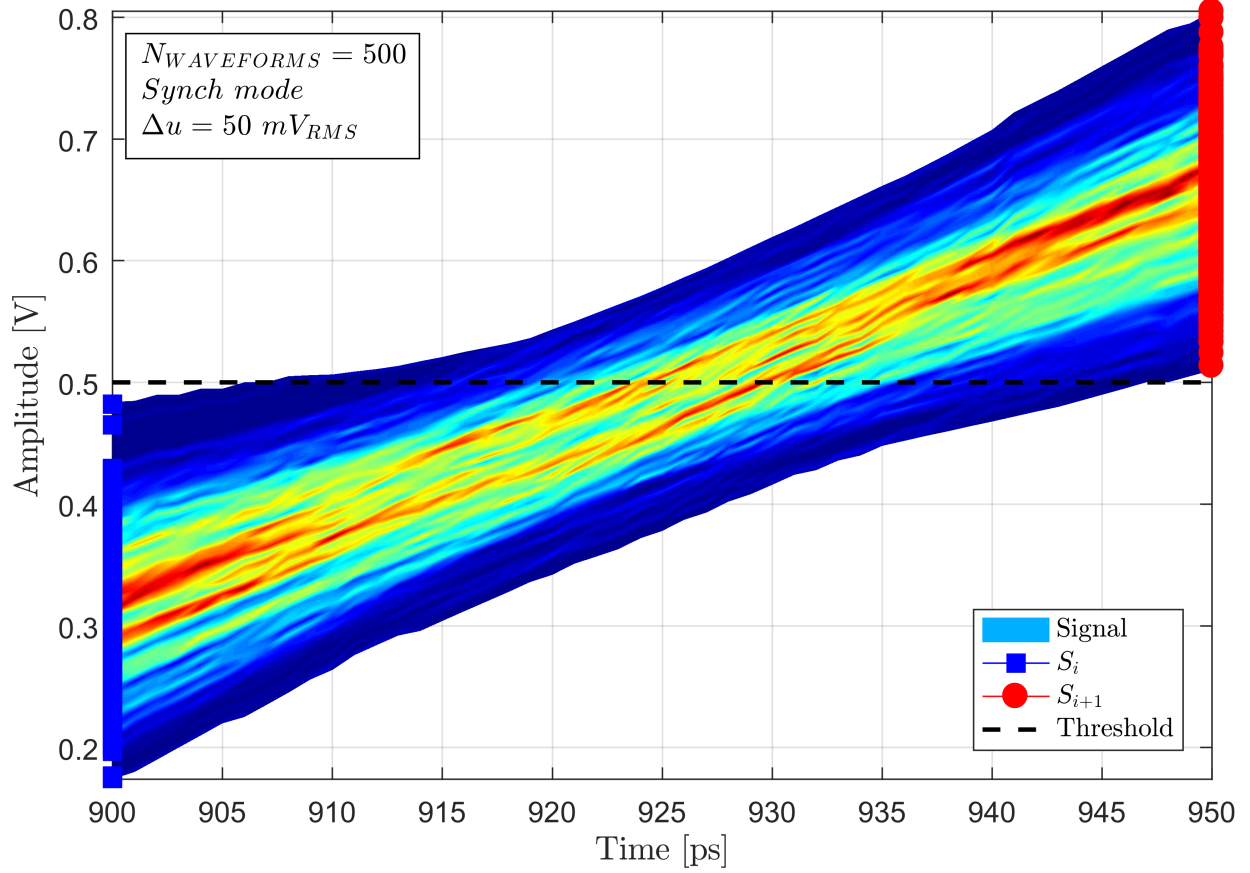


Figure 4.4: Persistence plot of 500 fitted waveforms over the transition point in the synchronous mode.

SIF values for a given statistical population used to determine $\sigma_{t_{INT.}}$. In this paper, this will be either the number of acquisition windows ($N_{acq.w.}$) or the number of periods within an acquisition window ($N_{periods}$). When the time of arrival of pulses is simulated or measured, $N_{acq.w.}$ is used. When the period (T_{PERIOD}) of a periodic signal is simulated or measured, $N_{periods}$ is used. Figure 4.5 shows $\sigma_{t_{INT.}}$ as a function of μ_{SIF} for $N_{acq.w.} = 1000$. The data shown has been simulated by using the square pulse shape in the synchronous mode and by sweeping the phase of both time bases. Results indicate that the resolution at $SIF = 0\%$ is better by a factor of approximately $\sqrt{2}$.

In order to consider the other effects separately, further case studies focus on the square pulse with the SWG operating in the synchronous mode.

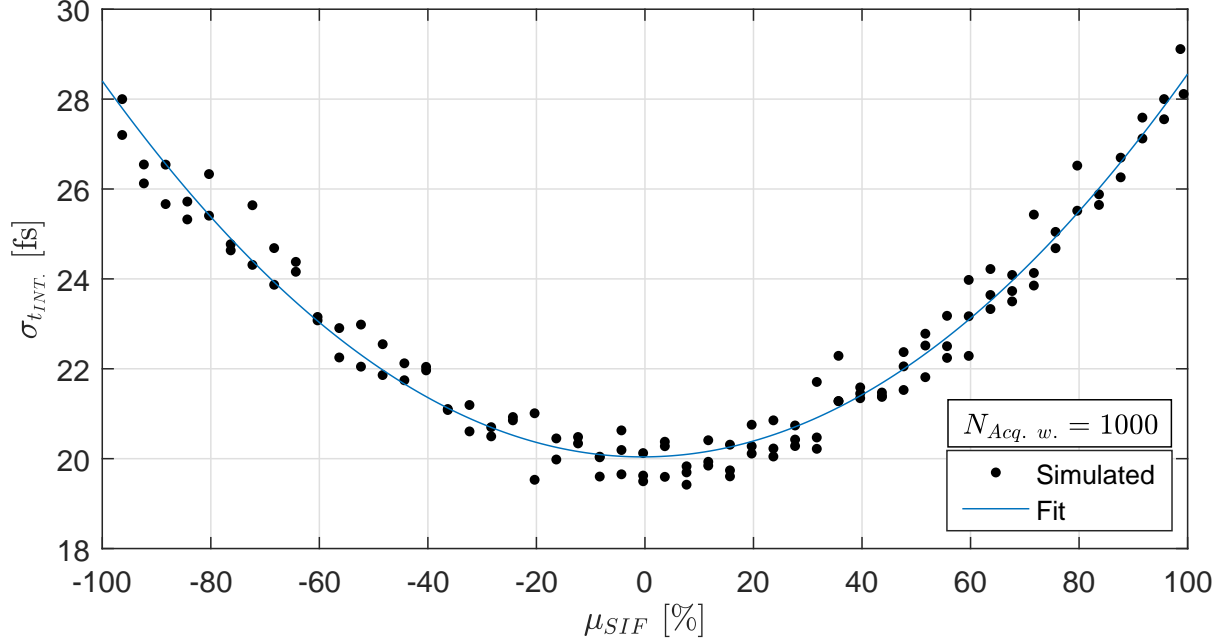


Figure 4.5: Interpolated transition time resolution as a function of the mean value of the sampling interpolation fraction.

4.4.2 Amplitude Noise vs Quantization Effects

Amplitude noise (Δu_{AMP}) is added to the signal as a random, normally distributed disturbance of the signal amplitude. Quantization of sampled values is implemented by simply rounding the values with a resolution equal to one half of the least significant bit (LSB) and is therefore uniformly distributed. The quantization error has been accounted for in the theoretical estimates by adding a quantization noise (Δu_Q) term to the total RMS noise as shown in Equation 4.3 [22]. Figure 4.6a shows a simulated and estimated $\sigma_{t_{INT.}}$ as a function of N_{BITS} .

$$\Delta u = \sqrt{\Delta u_{AMP}^2 + \Delta u_Q^2} \quad (4.3)$$

Estimated and simulated values diverge significantly at low N_{BITS} . Figure 4.6b shows the signal spectrum sampled at three different N_{BITS} . At low N_{BITS} , quantization distorts the signal, thus giving rise to higher harmonics. Consequently, quantization noise is an insufficient way of modeling the quantization error.

Figure 4.7a shows histograms of the simulated $t_{INT.}$ for the cases of $N_{BITS} = 8$ (left) and $N_{BITS} = 12$ (right), respectively. The 8-bit case histogram shows two peaks indicating that the $t_{INT.}$ distribution is dominated by the uniformly distributed quantization error. Conversely, the 12-bit case histogram shows a normal distribution, which can be associated with amplitude noise. Sweeping $\sigma_{t_{INT.}}$ as a function of Δu and N_{BITS} , as shown in Figure 4.7b, reveals that quantization

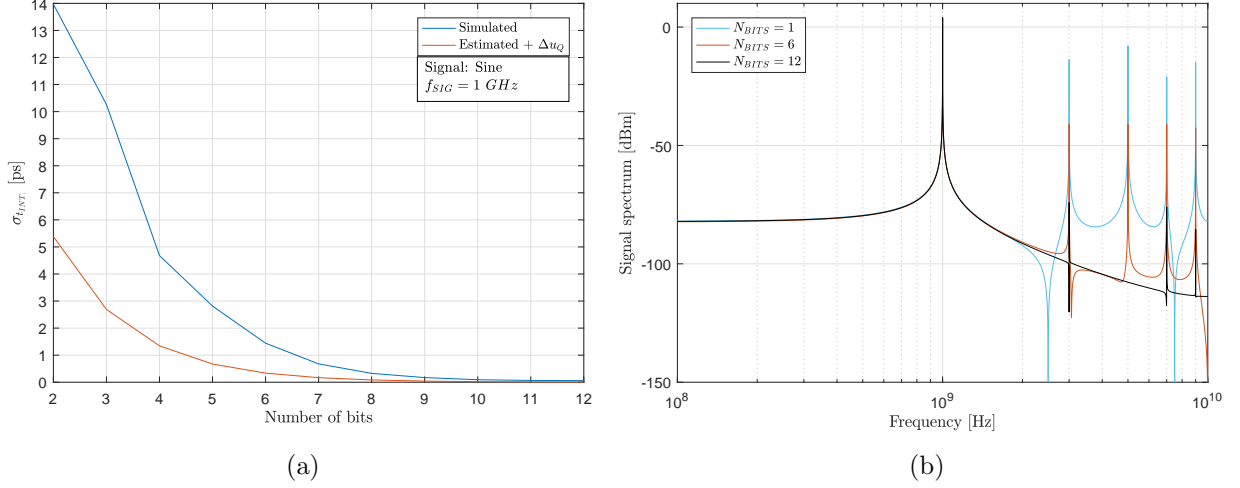


Figure 4.6: (a) Simulated and estimated interpolated transition time resolution as a function of N_{BITS} . (b) Signal spectrum for sampled signals for three different vertical resolutions in terms of N_{BITS} .

effects take over when the N_{BITS} falls below 9.5 for the given values of f_{SAMP} and BW . It should be noted that this transition is not easily observed in the asynchronous mode.

Another important aspect to consider is the effect of the decreasing signal amplitude on the timing resolution. Figure 4.8 shows the interpolated transition time resolution along with the corresponding μ_{SIF} as functions of the signal amplitude. At very low amplitudes, the simulation and estimated values match. However, at larger amplitudes, amplitude noise starts to dominate, at which point the SIF value becomes important.

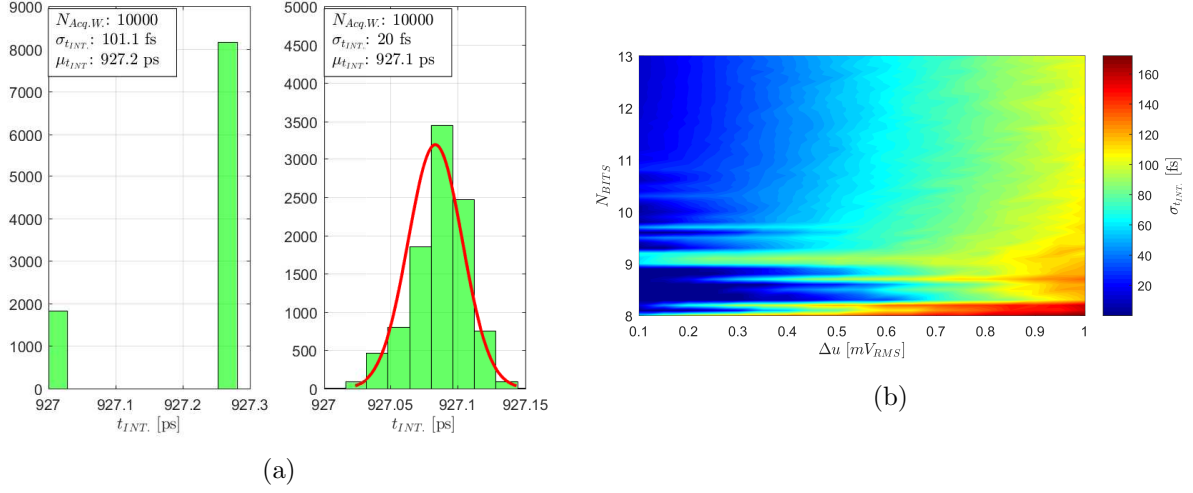


Figure 4.7: (a) Histograms of simulated interpolated transition times for the 8-bit (left) and 12-bit (right) resolution cases respectfully. (b) Interpolated transition time resolution as a function of noise and N_{BITS} .

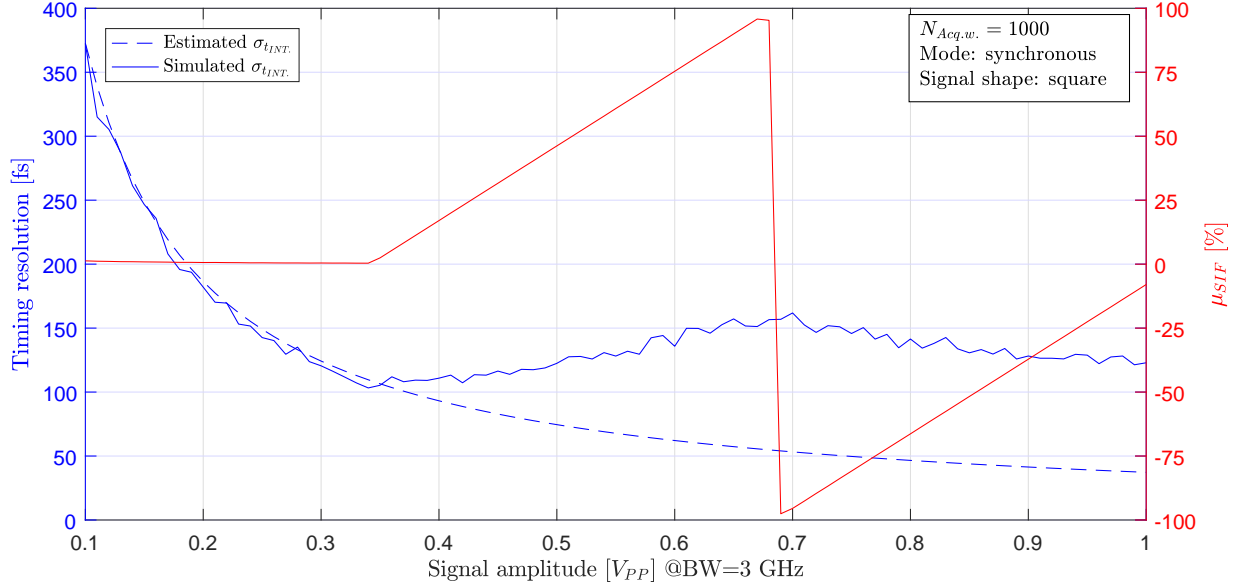


Figure 4.8: Interpolated transition time resolution along with the corresponding μ_{SIF} as functions of signal amplitude. Each signal voltage point has been simulated with a population of 1000.

4.4.3 Jitter

Jitter is injected into the sampling time base by adding a random, normally distributed time offset to the sampling time base vector.

The effect of jitter in Equation 4.1 is accounted for in the amplitudes of the samples. The calculations predict a timing resolution increase by the amount equal to the injected jitter multiplied by $\sqrt{2}$. Simulation results confirm this prediction with good agreement between the calculated (82.7 fs) and simulated (84.1 fs) values. Figure 4.9 shows $\sigma_{t_{INT.}}$ as a function of added jitter (J_{ADDED}) for N_{BITS} ranging from eight to twelve. The added jitter starts to dominate above $\approx 100 \text{ fs}$ and it completely dominates $\sigma_{t_{INT.}}$ in the picosecond regime.

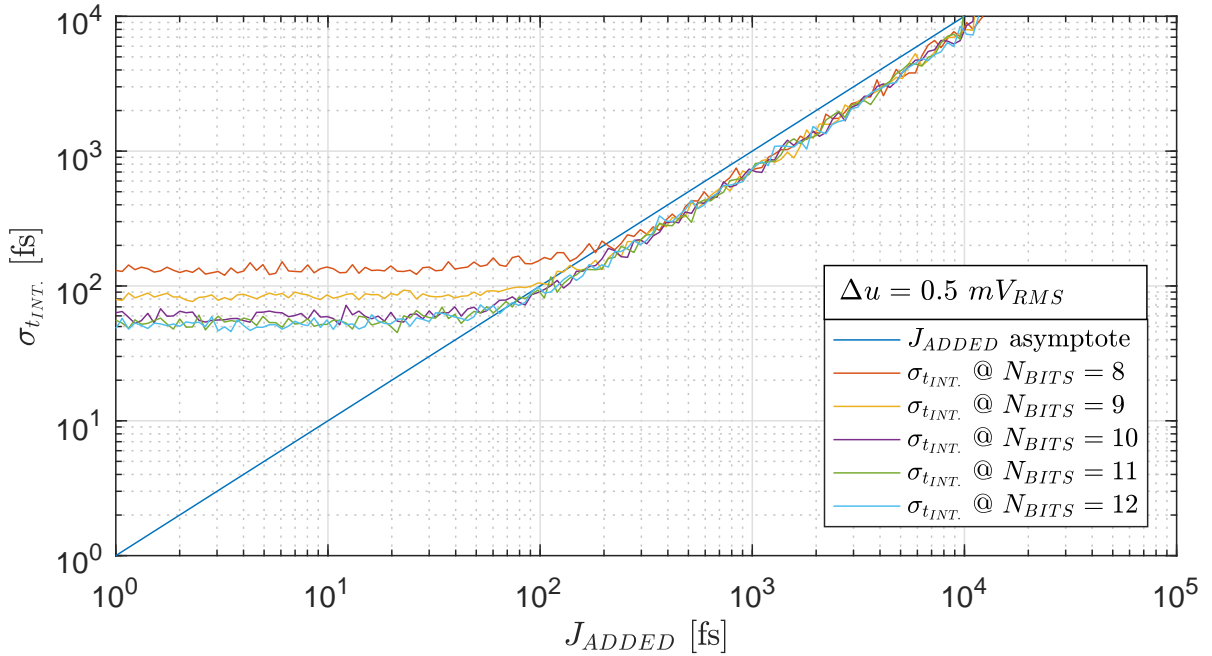


Figure 4.9: Interpolated transition time resolution as a function of added jitter at different numbers of bits.

4.4.4 Scaling of $\sigma_{t_{INT.}}$ as a Function of Sampling Frequency, Bandwidth, and Noise

Upon further inspection of Equation 2.8, we observe that it allows for a significant decrease in timing resolution by simply manipulating one of the parameters, while the other two remain constant. However, limits to this scaling are intuitively expected. Parameter sweeps of the $\sigma_{t_{INT.}}$ as a functions of $f_{SAMP.}$, BW , and Δu reveal some of these limits.

Figure 4.10 shows the simulated interpolated transition time resolution as a function of noise and sampling speed along with the corresponding values of μ_{SIF} .

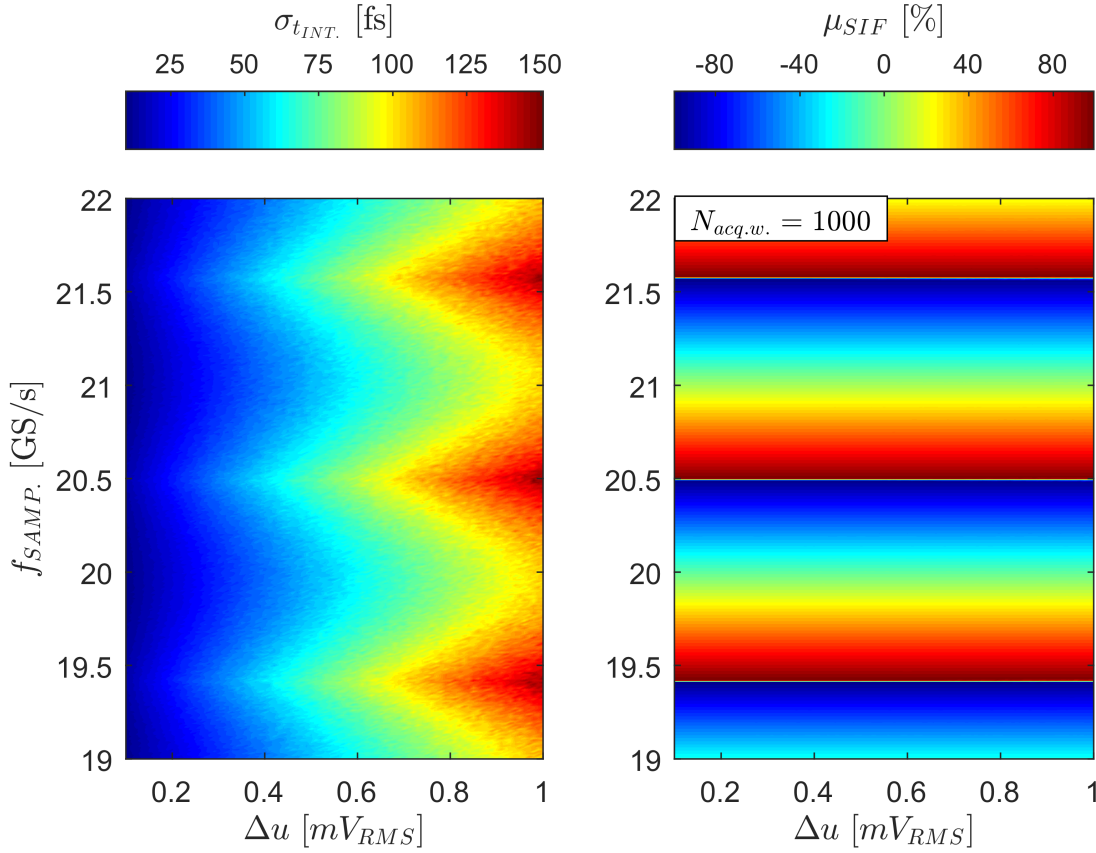


Figure 4.10: Simulated interpolated transition time resolution (left) and the corresponding μ_{SIF} (right) as functions of noise and sampling speed.

Results indicate that $\sigma_{t_{INT.}}$ periodically improves and deteriorates at specific intervals of the $f_{SAMP.}$ values. These changes in resolution are directly correlated with the μ_{SIF} . As expected, the lowest resolution occurs when the μ_{SIF} is closest to 0 %. The same effect can be observed by sweeping the bandwidth or the rise time (t_r) of the signal. The t_r of the pulse is directly related to the analog bandwidth through the relation $t_r \approx 0.35/BW$ [57]; higher bandwidths lead to shorter t_r and thus narrower pulses. Figure 4.11 shows $\sigma_{t_{INT.}}$ as a function of bandwidth and sampling speed, where spots of low $\sigma_{t_{INT.}}$ can be clearly identified.

The low $\sigma_{t_{INT.}}$ spots occur where the apparent phase of the signal aligns with the sampler time base in a way that the SIF value is closest to 0 %. This effect is not the randomly varying phase of the asynchronous mode of operation; instead, the observed effect comes from an apparent rotation of the phase in time between the signal and the sampling time bases due to the fractional relationship of the signal frequency and the sampling speed. The effects of both these mechanisms are integrated in the SIF and thus become indistinguishable in the measurements.

In addition, decreasing the t_r of the pulse beyond a certain point causes the sampling period

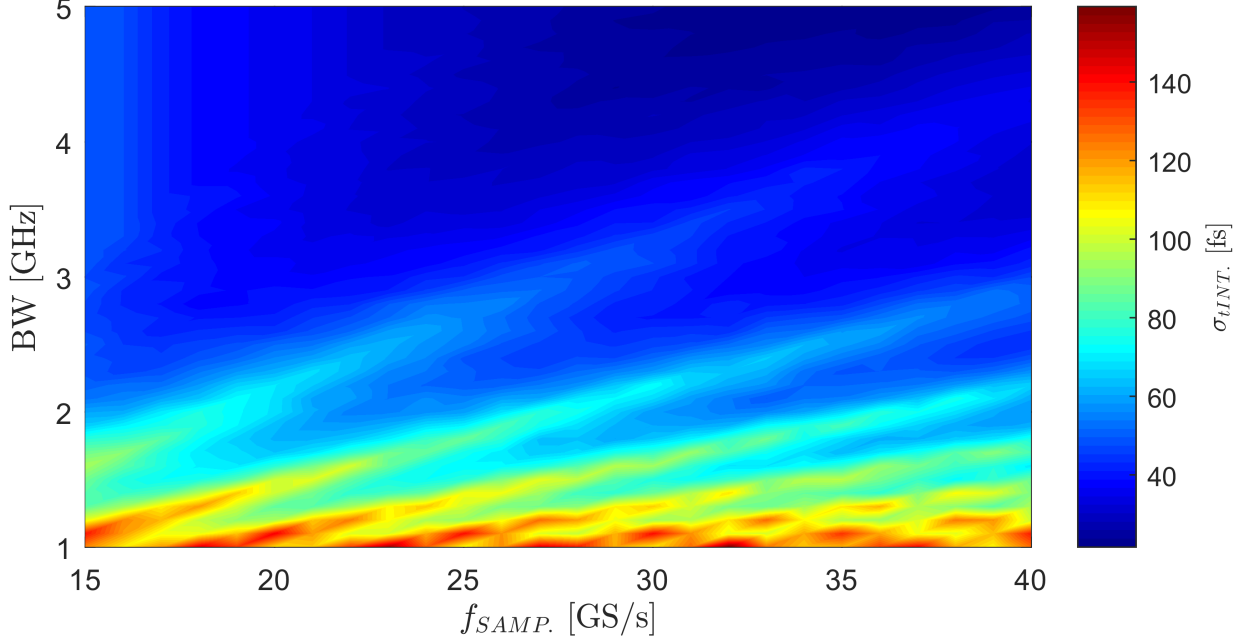


Figure 4.11: Interpolated transition time resolution as a function of sampling speed and bandwidth with spots of low $\sigma_{t_{INT.}}$.

to be wider than the t_r of the pulse. Consequently, one of the sampling points falls off the leading edge, thus considerably worsening the $\sigma_{t_{INT.}}$. For the sampling period of 50 ps (20 GS/s), the highest t_r is 80 ps, which corresponds to the analog bandwidth of 4.25 GHz. In our case, rise time is measured from 10 % to 90 % of the signal amplitude U .

Derivation:

1. The t_r and BW are connected through the relation given in Equation 4.4 [57].

$$t_r \approx \frac{0.34}{BW} \quad (4.4)$$

2. The slew rate (SR) can be calculated as shown in Equation 4.5.

$$SR = \frac{dV}{dt} = \frac{0.9 \cdot U - 0.1 \cdot U}{t_r} \quad (4.5)$$

3. In order for the sampling points not to fall off the leading edge, the maximum SR is given in Equation 4.6,

$$SR_{limit} = \frac{U}{2 \cdot T_{SAMP.}} \quad (4.6)$$

where $T_{SAMP.}$ is the sampling period, which is defined as: $T_{SAMP.} = 1/f_{SAMP.}$.

4. By equating SR and SR_{limit} and substituting in the introduced relations for t_r and $T_{SAMP.}$, we obtain Equation 4.7.

$$BW_{limit} \approx 0.21 \times f_{SAMP.} \quad (4.7)$$

Henceforth, for pulse-like waveforms, it can be postulated that, in order to get a good measurement of $\sigma_{t_{INT.}}$, the upper limit of the analog signal bandwidth for a given sampling speed can be estimated by using Equation 4.7. It is worth noting that this is lower than the bandwidth limitation imposed by the Nyquist theorem [58] for signal reconstruction.

Furthermore, $\sigma_{t_{INT.}}$ does not seem to improve with the increasing sampling speed as estimated by Equation 2.1. Figure 4.12 shows $\sigma_{t_{INT.}}$ in the limiting case as a function of $f_{SAMP.}$. After a certain point, the resolution is completely dominated by the noise through the AM-to-PM mechanism.

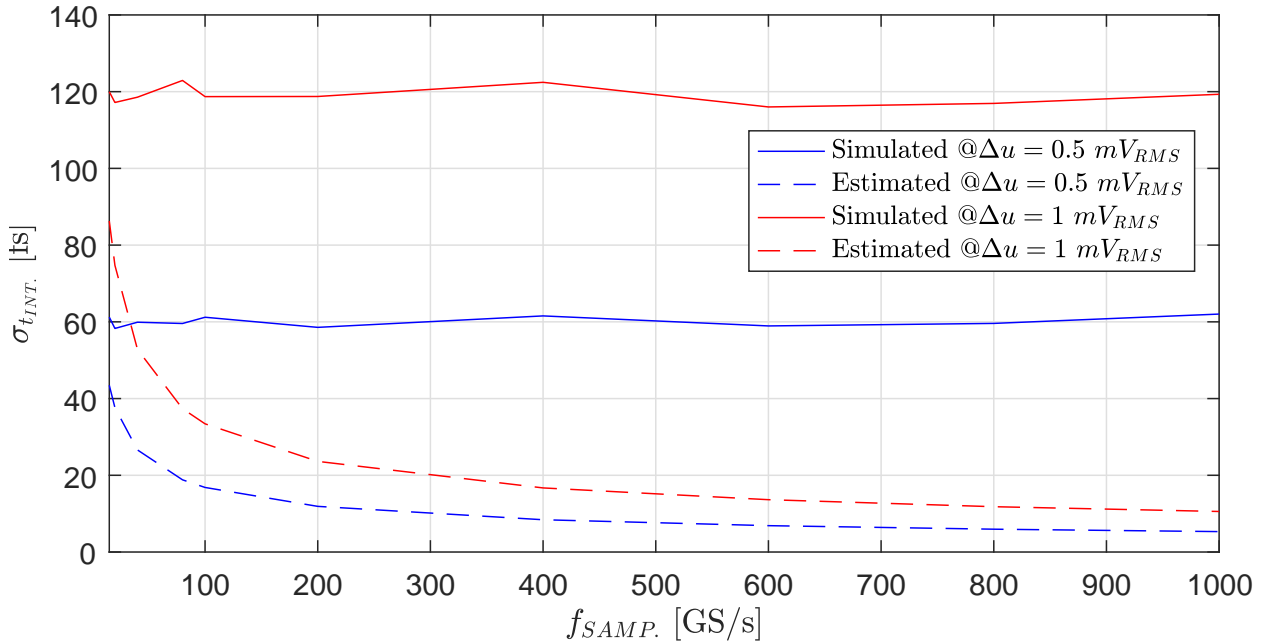


Figure 4.12: Asymptotic interpolated transition time resolution as a function of sampling speed.

4.4.5 Timing Extraction Algorithm Comparison

One source of error that is not hardware related is the choice of extraction algorithm. In this section, we compare different extraction algorithms on three different signal shapes: sine, square pulse, and Gaussian pulse. In all cases the signals are generated using the SWG with the parameters specified in Table 2.3 along with 500 fs of injected jitter ($J_{INJ.}$).

The extraction algorithms used are:

- Linear extraction algorithm represented by Equation 4.1;
- A template fit based on least-mean-square (LSQ) approximation [59];
- A monotone piecewise cubic interpolation technique [60] with the use of two MATLAB[®] functions: spline [61] and phcip [62].

Figure 4.13 shows the interpolated transition time resolution as a function of the number of samples for a sine wave signal. In case of the LSQ algorithm, the number of samples effects the timing performance considerably. That is, the timing ambiguity caused by jitter and apparent phase rotations is averaged out by the increasing number of samples. The other algorithms are not affected by the number of samples since they are all defined piece by piece, hence the term piecewise interpolation. However, they do provide different results for both modes of operation, suggesting a dependence on the SIF FOM.

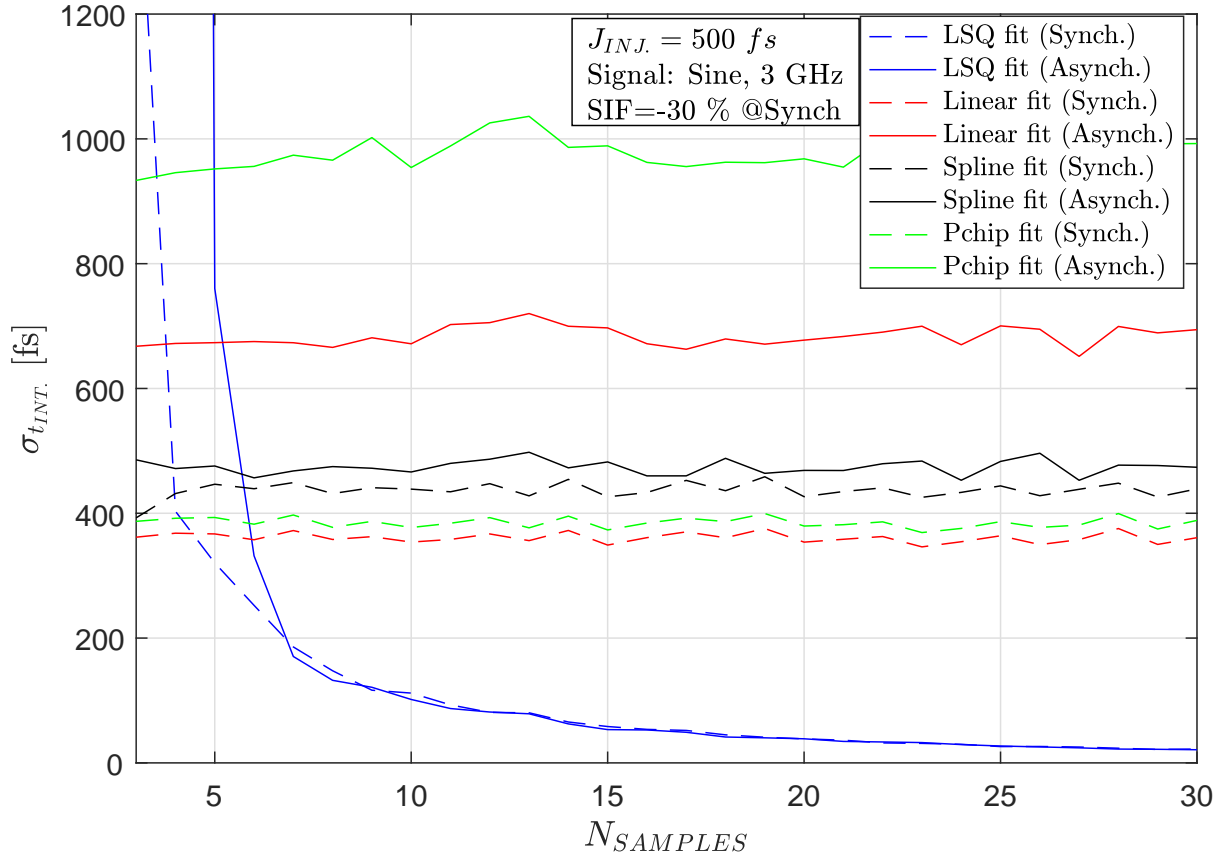


Figure 4.13: Timing resolution as a function of number of samples by using different extraction algorithms for both, synchronous and asynchronous cases.

Figure 4.14 shows the extracted timing resolution as a function of the SIF for a sine waveform with 40 samples for four extraction algorithms. In the case of the LSQ algorithm, the effect of the SIF is negligible on the corresponding timing resolution. The linear extraction algorithm shows the already discussed parabolic dependence on the SIF. The piecewise approximation algorithms show an asymmetric dependence on the SIF, which is caused by the slope transition from concave to convex at the threshold. That is, the cubic approximation uses polynomials, which have convex slopes, thus providing a better fit on convex slopes.

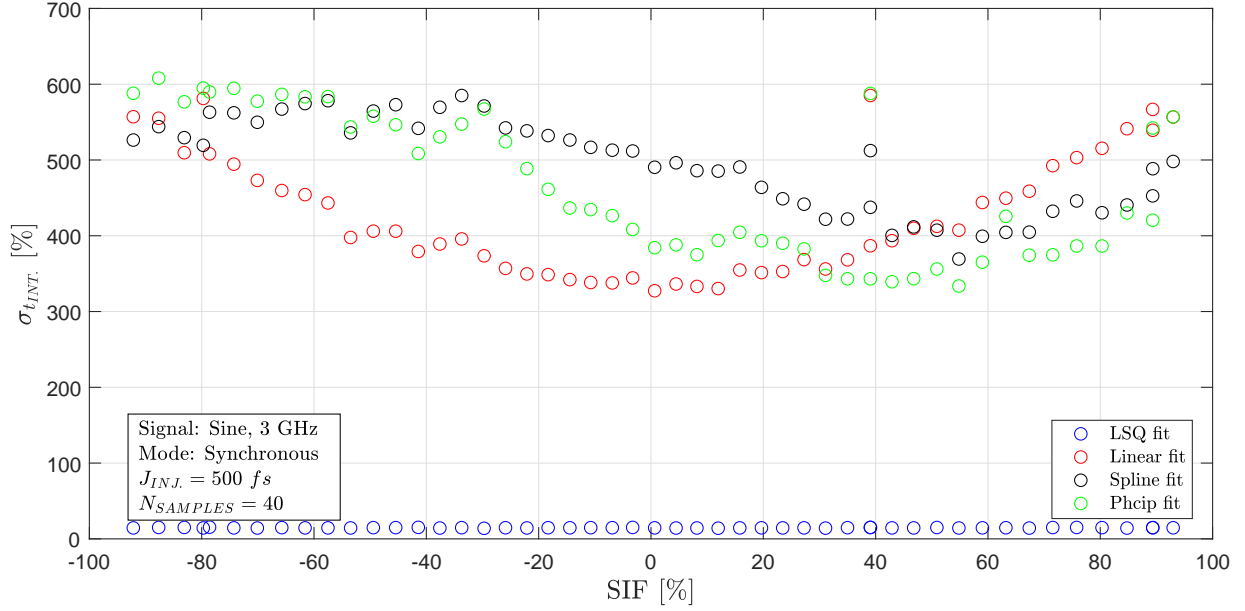


Figure 4.14: Timing resolution as a function of the SIF by using different extraction algorithms (synchronous mode) on a sine waveform.

Figures 4.15a and 4.15b show the extracted timing resolution as a function of the SIF for a square and a Gaussian pulse respectively. In this two cases, the LSQ algorithm is not considered.

The spline fit shows a very small dependence on the SIF for the square pulse waveform, while in the Gaussian pulse case, the dependence follows a quasisymmetric parabola. Conversely, the phcip fit exhibits a quasisymmetric parabolic SIF dependence for both cases.

As expected, the LSQ algorithm has proven to be superior in both modes of operation. The other three algorithms show similar performances in the synchronous mode of operation. Furthermore, if the SIF is close to zero, the linear fit seems to be the best choice. In the asynchronous mode of operation, the spline fit shows the best performance.

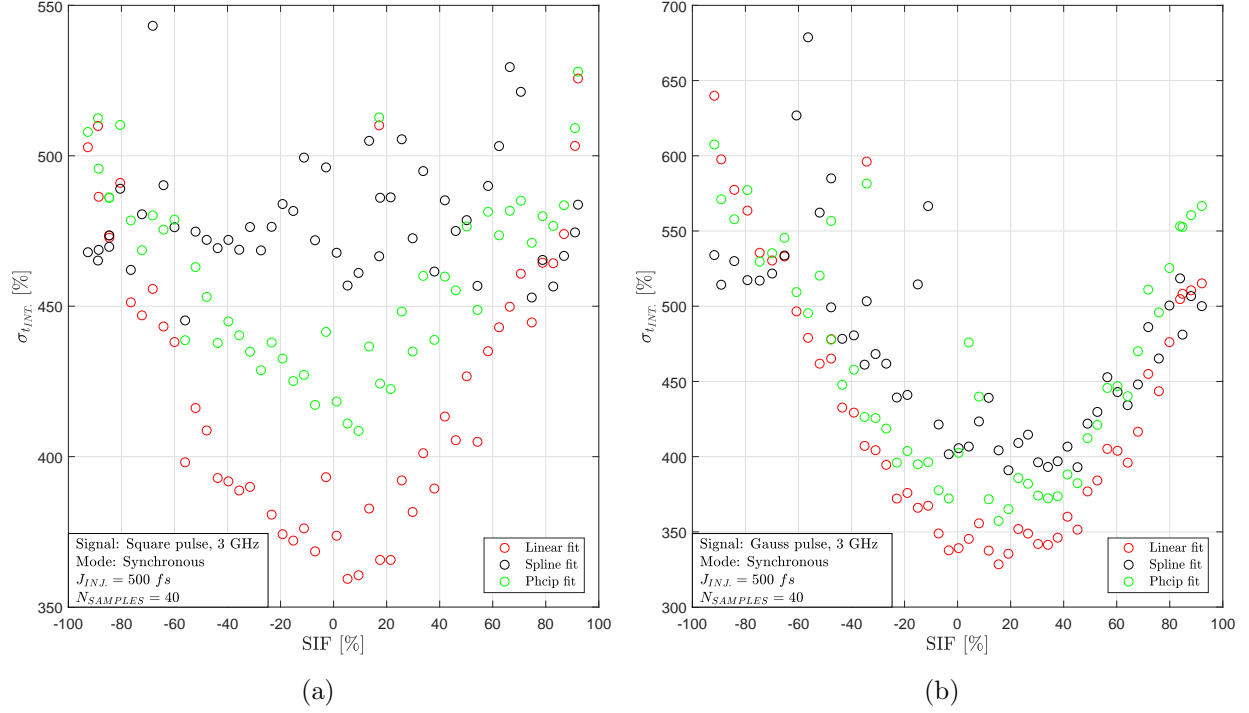


Figure 4.15: (a) Timing resolution as a function of the SIF by using different extraction algorithms (synchronous mode) on a square pulse waveform. (b) Timing resolution as a function of the SIF by using different extraction algorithms (synchronous mode) on a Gaussian pulse waveform.

4.4.6 Summary of Simulation Results

Simulation results indicate that there are four main sources of error with the associated mechanisms that significantly impact the timing resolution at the femtosecond scale. The list below summarizes these four main sources of error:

- Choice of extraction algorithm;
- Jitter;
- Noise;
- Distortion.

The first item on the list is an effect that is purely software-dependent. As shown in 4.4.5, using a curve-fitting algorithm with a LSQ approximation technique for minimizing the error can yield significantly better results.

From the hardware perspective, jitter, noise, and distortion are the main sources of error that degrade the timing performance through various mechanisms. These mechanisms are summarized below:

- AM-to-PM conversion;
- Quantization;
- Random time-varying phase difference between the signal and the sampling time bases in the asynchronous mode of operation;
- Apparent periodic rotation of the phase in time between the signal and the sampling time bases due to the fractional relationship of the signal frequency ($f_{SIG.}$) and sampling speed.

The dominant mechanism that converts amplitude noise into phase noise is the AM-to-PM conversion. In time domain it manifests as jitter. Quantization gives rise to distortion, which can be quantified by the total harmonic distortion (THD) FOM. Depending on the operating mode, the other two mechanisms can also have a significant impact on the $\sigma_{t_{INT.}}$. Both are quantified by the SIF FOM.

4.5 Test Bench Measurements

4.5.1 Test Bench Setup

In order to verify the validity of the SWG, a test bench was set up as described below.

Unfortunately, only a sine wave signal from a high-performance RF generator had the necessary low jitter to perform these tests. The timing resolution was tested so that the period T_{PERIOD} of the signal was extracted using the linear fit as defined in Equation 4.1, where the standard deviation of the cycle to cycle period is taken as the resolution of the extracted signal period ($\sigma_{T_{PERIOD}}$) in a given acquisition window. Figure 4.16 shows how the period and the corresponding SIF are extracted. In this case, the period is defined by two interpolated zero-crossings on positive slopes. When crossing the zero point, the sine waveform has the most linear-like response and, at the same time, the signal rate of change is the highest, thus limiting the AM-to-PM conversion, leading to better timing results. Two consecutive periods are defined by three points, thus using only one value of the SIF is insufficient. For this reason, the mean of all SIF values (μ_{SIF}) within an acquisition window is used for plotting the results.

Three oscilloscopes (OSCs) and one ADC have been used to provide data acquisition (DAQ). Figure 4.17 shows the block diagram of the test bench setup.

Equipment used in the test is summarized in Table 4.2.

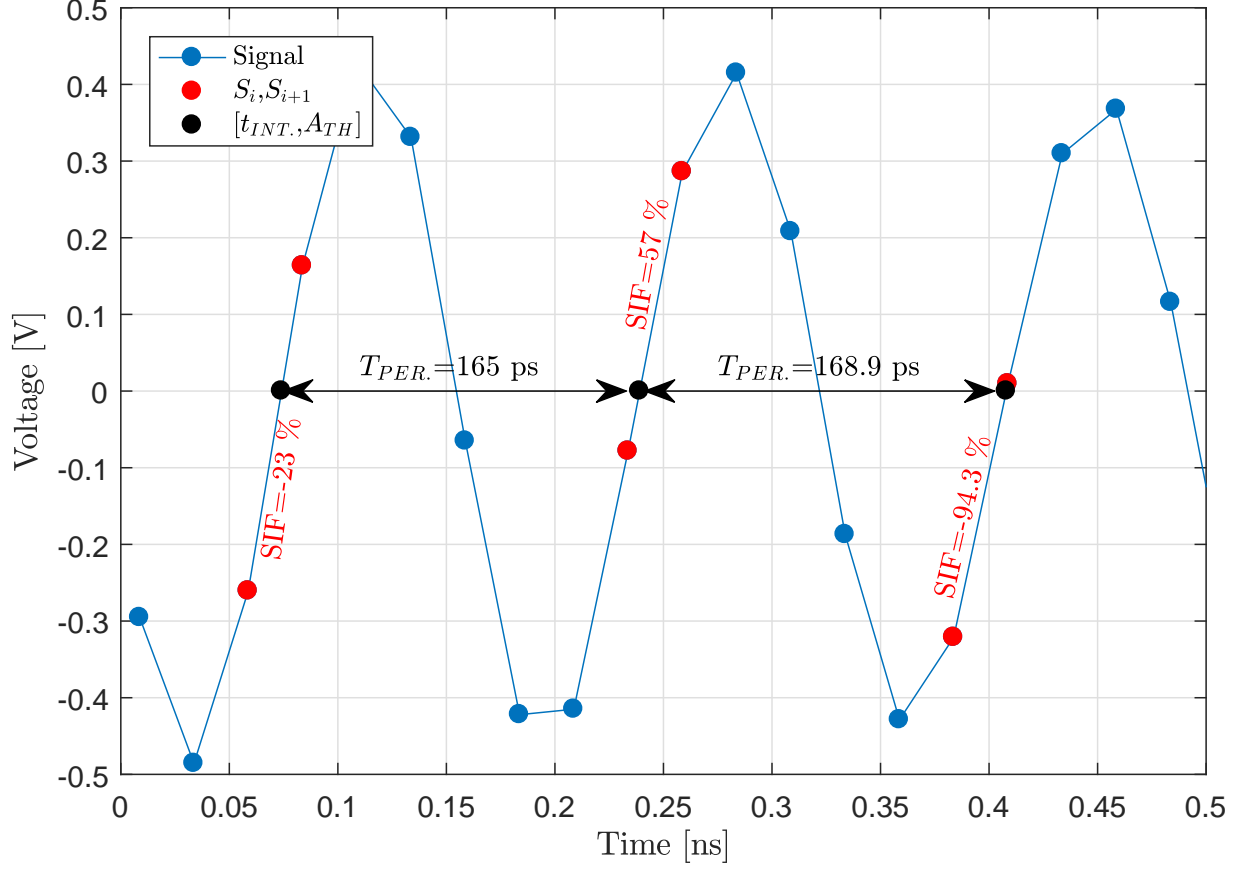


Figure 4.16: Signal period extraction in time domain with the corresponding SIF.

Table 4.2: Test bench equipment.

<i>Equipment</i>	<i>Unit description</i>
RF generator	Agilent E4428C [63]
Splitter	Mini-circuit ZFSC-2-2500
ADC	Ti EVMADC12J4000 [64]
Oscilloscope 1 (OSC1)	LeCroy SDA13000 [65]
Oscilloscope 2 (OSC2)	Tektronix TDS6804B [66]
Oscilloscope 3 (OSC3)	Agilent MSO8104A [67]
Computer	Dell XPS 8700
RF Cables	Mini-circuit 1021 141-6SM+

Performance parameters BW , $f_{SAMP.}$, J_{ADDED} , Δu , and N_{BITS} are summarized in Table 4.3.

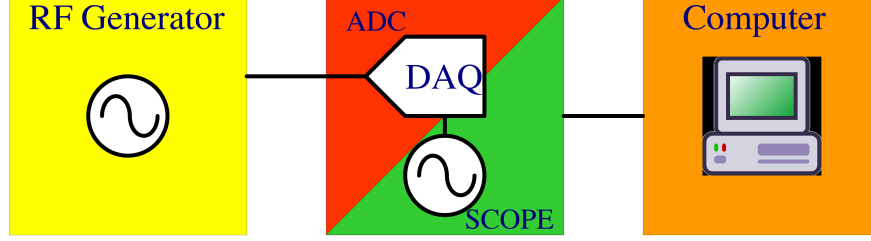


Figure 4.17: Test bench setup block diagram (asynchronous mode).

Table 4.3: Test bench equipment performance parameters.

<i>Equipment</i>	<i>BW</i>	<i>f_{SAMP.}</i>	<i>J_{ADDED}</i>	<i>Δu</i>	<i>N_{BITS}</i>
Agilent E4428C [63]	$\leq 6 \text{ GHz}$	NA	$\approx 74 \text{ fs}$	NA	NA
Ti EVMADC12J4000 [64]	$(0.4 - 3) \text{ GHz}$	$\leq 4 \text{ GS/s}$	$\approx 485 \text{ fs}$	0.33 mV_{RMS}	12
LeCroy SDA13000 [65]	$DC - 13 \text{ GHz}$	40 GS/s	2.5 ps Trigger	0.64 mV_{RMS}	8
Tektronix TDS6804B [66]	$DC - 8 \text{ GHz}$	20 GS/s	1.5 ps Trigger	0.61 mV_{RMS}	8
Tektronix MSO8104A [67]	$DC - 1 \text{ GHz}$	4 GS/s	5 ps Period	0.50 mV_{RMS}	8

4.5.2 Measurement Results

Two sets of measurements were conducted. In the first set only oscilloscopes were used as the DAQ devices with the vertical resolution limited to eight bits and an unknown inner structure. The added jitter information was retrieved from their respective data sheets, while the baseband noise floors were measured for the individual oscilloscopes at the properly terminated (50Ω) inputs.

The second set of measurements was performed using a Texas Instruments ADC12J4000 ADC [68]. The ADC comes on an evaluation board along with all of the necessary documentation (schematic and board layout files), thus providing a good estimate of the added jitter. Furthermore, the evaluation board has an easy-to-use graphical user interface (GUI), which allows for an easy reconfiguration of the ADC and the semi-automation of the tests at the same time.

In all cases only raw sampled data with no interpolation or filtering was considered. The analysis was done offline using MATLAB[®] software.

Oscilloscope Measurement Results

Table 4.4 shows the oscilloscope measurements results along with the corresponding simulated results from the SWG configured using the data in Table 4.3.

Table 4.4: Summary and comparison of oscilloscope measurements and SWG simulations results.

<i>DAQ</i>	<i>f_{SIG.}</i>	<i>Measured $\sigma_{T_{PERIOD}}$</i>	<i>Simulated $\sigma_{T_{PERIOD}}$</i>
OSC1	3 <i>GHz</i>	1.78 <i>ps</i>	2.96 <i>ps</i>
OSC1	6 <i>GHz</i>	2.28 <i>ps</i>	3.03 <i>ps</i>
OSC2	3 <i>GHz</i>	1.38 <i>ps</i>	1.42 <i>ps</i>
OSC2	4 <i>GHz</i>	1.68 <i>ps</i>	1.58 <i>ps</i>
OSC3*	400 <i>MHz</i>	5.68 <i>ps</i>	5.86 <i>ps</i>
OSC3*	500 <i>MHz</i>	4.27 <i>ps</i>	5.75 <i>ps</i>

* Signal amplitude equal to 0.475 *Vpp*.

Figure 4.18a shows a comparison between the OSC1 test and the corresponding simulation data in terms of the $\sigma_{T_{PERIOD}}$ as a function of μ_{SIF} . There are two things to note:

1. The analysis on the simulated data reports a worse period resolution than the one reported for the OSC1 data;
2. The spread of the data points nonuniform across the μ_{SIF} space, which suggests a non uniform phase distribution.

Figure 4.18b shows a comparison of the measured and simulated data in terms of the distributions of $\sigma_{T_{PERIOD}}$, μ_{SIF} , and σ_{SIF} . The μ_{SIF} distributions are almost identical, while the $\sigma_{T_{PERIOD}}$ distributions indicate that the $\sigma_{T_{PERIOD}}$ of the OSC1 data exhibits a narrower, but still normal distribution. By matching the SWG to the OSC1 data and comparing the results with the measurements at $f_{SIG.} = 3$ *GHz* and $f_{SIG.} = 6$ *GHz*, it has been determined that the reported jitter in the OSC1 data sheet was inaccurate. We believe the actual added jitter to be approximately 1.9 *ps*.

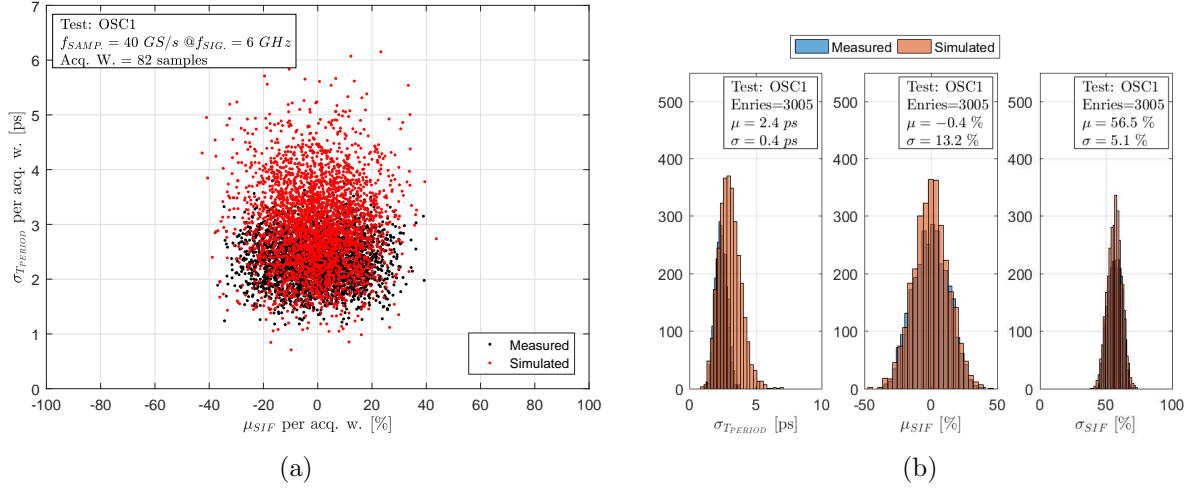


Figure 4.18: (a) Comparison between the OSC1 test data and the corresponding simulation in terms of the extracted signal period resolution as a function of μ_{SIF} . (b) Distributions of the extracted signal period resolution (left), μ_{SIF} (center), and σ_{SIF} (right) for the OSC1 test and the corresponding simulation with $f_{SIG} = 6 \text{ GHz}$.

Figure 4.19 shows a uniform distribution of the SIF. However, by observing the μ_{SIF} and σ_{SIF} distributions in Figure 4.18b, we notice that each μ_{SIF} value has a relatively large standard deviation. Figure 4.20 shows time domain plots of the measured data at $f_{SIG} = 6 \text{ GHz}$ and a simulation (with correct jitter input) at $f_{SIG} = 5.143 \text{ GHz}$. We notice that there is an apparent rotation of the phase at $f_{SIG} = 6 \text{ GHz}$, but not at $f_{SIG} = 5.143 \text{ GHz}$. As shown in the section 4.4.4, this is the consequence of the fractional relation between the (f_{SAMP}) and the (f_{SIG}). In the case of $f_{SIG} = 6 \text{ GHz}$, this ratio is 6.667. In contrast, at $f_{SIG} = 5.143 \text{ GHz}$ this ratio is 7, thus there is no apparent rotation of the phase. Due to this effect, the SIF values change considerably within one acquisition window, however all SIF values within the sampling interval have the same probability. Therefore, a statistical observation of the SIF values within this acquisition window yields a μ_{SIF} that is closer to the set threshold and a large σ_{SIF} . This explains their respective distributions seen in Figure 4.18b.

In the light of this, we conclude that $\sigma_{TPERIOD}$ is dominated by the added jitter, while the μ_{SIF} spread is dominated by the apparent rotation of the phase due to the fractional ratio between the sampling speed and the signal frequency at $f_{SIG} = 6 \text{ GHz}$ and $f_{SIG} = 3 \text{ GHz}$.

The OSC2 measurements and the corresponding simulation data match with good agreement.

The OSC3 measurements show good agreement with the simulation data at $f_{SIG} = 400 \text{ MHz}$, but not at $f_{SIG} = 500 \text{ MHz}$. Figure 4.21 shows the comparison between the OSC3 data and the corresponding simulation in terms of ($\sigma_{TPERIOD}$) as a function of μ_{SIF} . At first, it would seem that that the injected jitter is high, but tests and simulations with $f_{SIG} = 400 \text{ MHz}$ do not support such a conclusion. Therefore, further investigation is required.

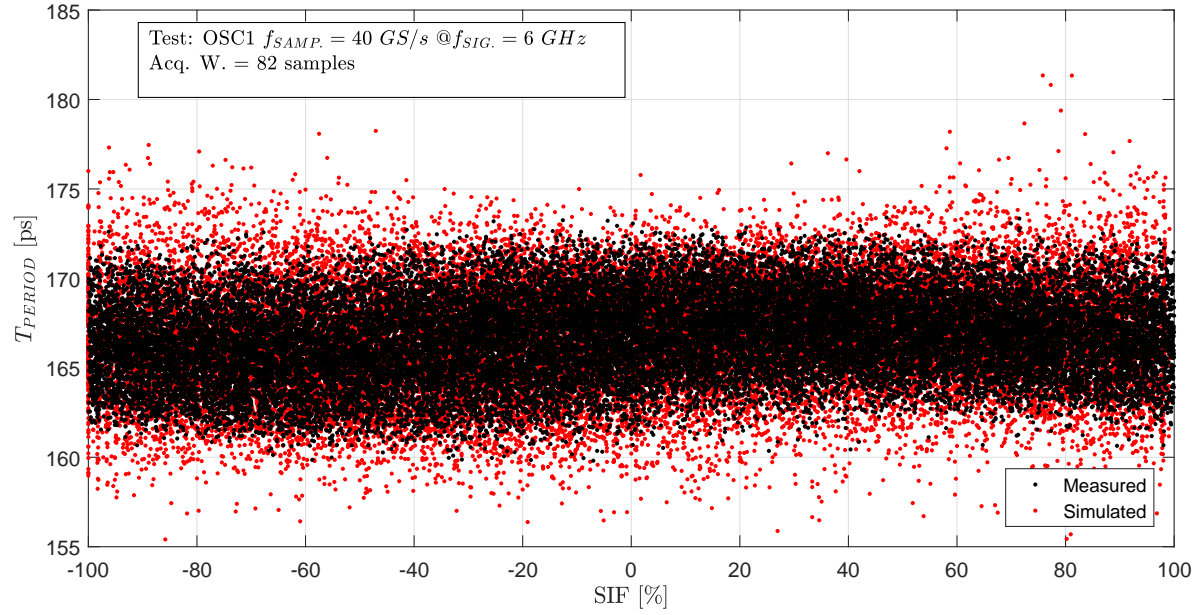


Figure 4.19: Comparison between the OSC1 test data and the corresponding simulation in terms of the extracted signal period resolution as a function of SIF.

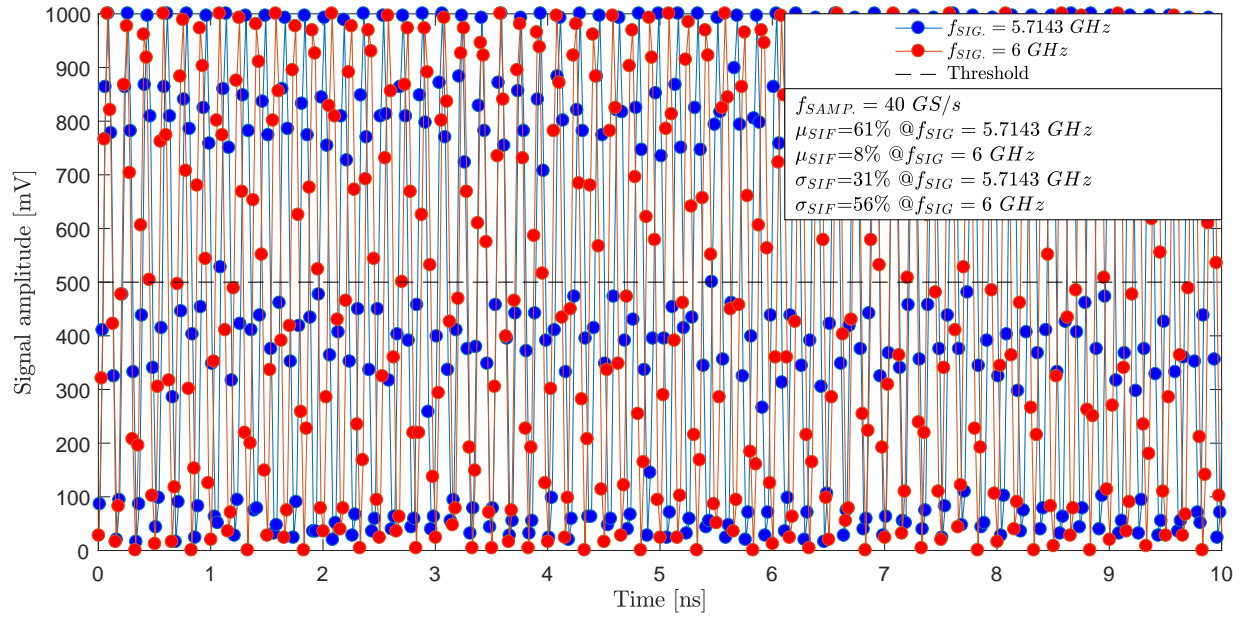


Figure 4.20: Sampled signal waveform in time domain at signal frequencies of 5.7143 GHz ($f_{SAMP.}/f_{SIG.} = 7$) and 6 GHz ($f_{SAMP.}/f_{SIG.} = 6.667$).

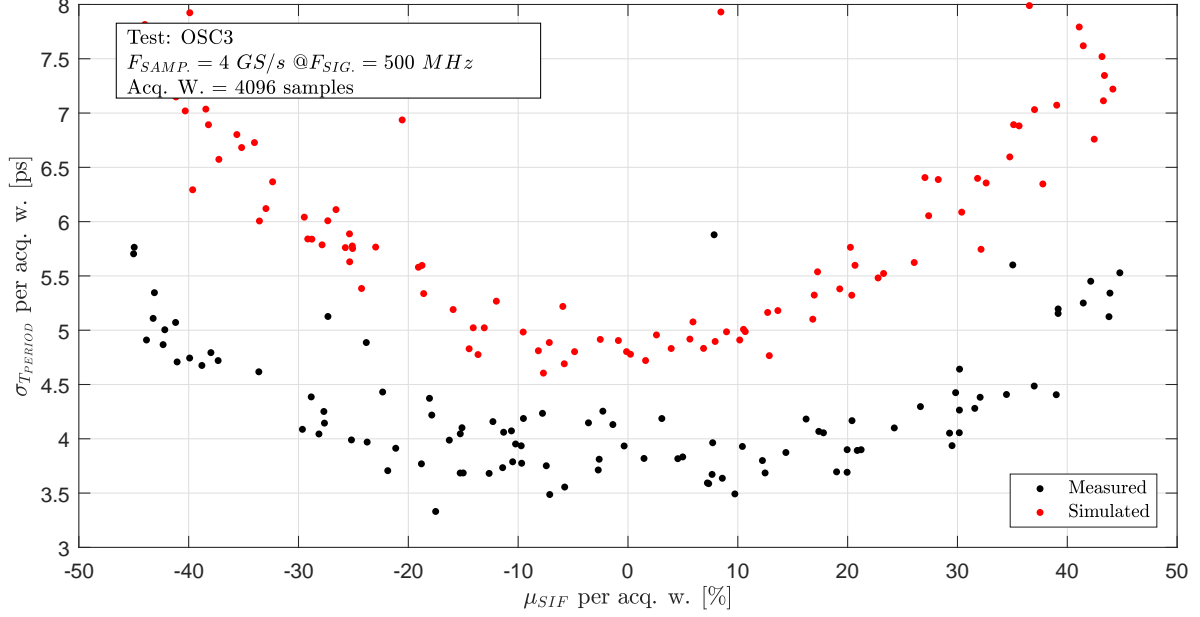


Figure 4.21: Comparison between the OSC3 test data and the corresponding simulation in terms of the extracted signal period resolution as a function of μ_{SIF} .

Also, it is worth noting that the μ_{SIF} spread is uniform, which is expected since $(f_{SAMP.}/f_{SIG.} = 8)$, thus no apparent rotation of the phase is present.

ADC Measurement Results (Asynchronous Mode)

Figure 4.22a shows the comparison between ADC measurements results and the corresponding simulation in terms of $\sigma_{TPERIOD}$ as a function of μ_{SIF} at a sampling speed of 4 GS/s with an input signal frequency of 500 MHz. The measured and the simulated results show very good agreement.

With the ability to change the ADC sampling speed, a parameter sweep has been performed in terms of $\sigma_{TPERIOD}$ as a function of available values of $f_{SAMP.}$ and $(f_{SIG.})$. Figure 4.22b shows the results. The patterns of low resolution spots, as seen in Figure 4.11, can be identified in the measured data, thus confirming the predictions from the simulation.

Figures 4.23a and 4.23b show the $\sigma_{TPERIOD}$ as a function of the μ_{SIF} at different sampling speeds and signal frequencies, respectively.

As expected, the μ_{SIF} spread is lower at fractional values of the $f_{SAMP.}/f_{SIG.}$ ratio. The $\sigma_{TPERIOD}$ is dominated by this effect due to low added jitter. Consequently, the μ_{SIF} is non-uniform and the $\sigma_{TPERIOD}$ is large, even though the μ_{SIF} per acquisition window is small. As noted before, the SIF standard deviation σ_{SIF} per acquisition window is large for these datasets. Therefore, to better visualize this effect, the RSD or coefficient of variation (CV) [69] of the SIF parameter (RSD_{SIF}) can be used. The RSD is defined as the ratio of the standard deviation

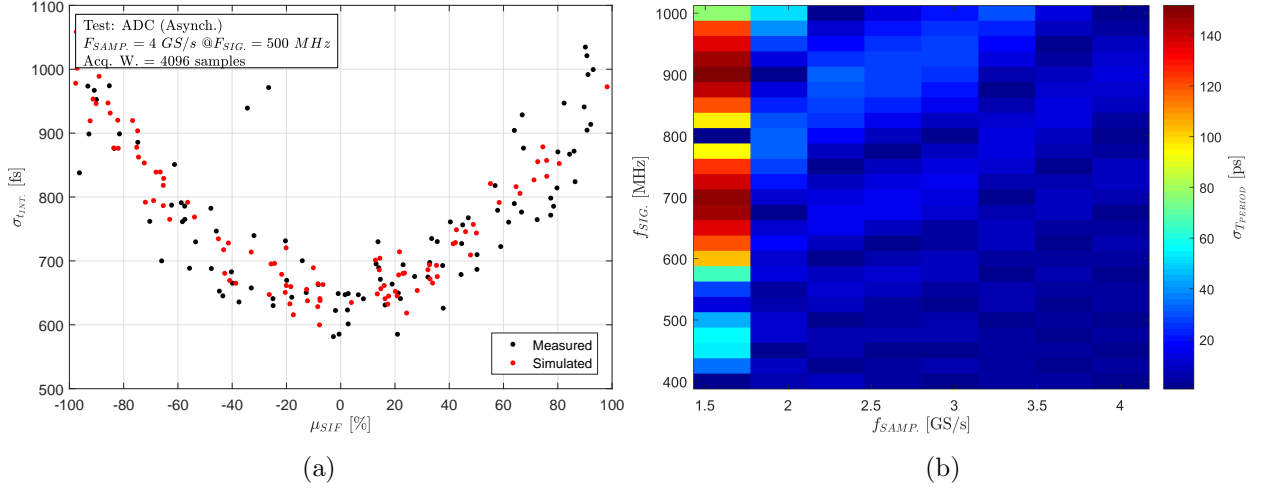


Figure 4.22: (a) Comparison between ADC test data and the corresponding simulation in terms of the extracted signal period resolution as a function of μ_{SIF} . (b) Extracted signal period resolution from the ADC data as a function of f_{SAMP} and f_{SIG} .

divided by the mean. In other words, it shows how large the spread is in comparison to the mean of a dataset. Equation 4.8 defines the RSD.

$$RSD_{SIF} = \frac{\sigma_{SIF}}{\mu_{SIF}} [\%] \quad (4.8)$$

Figures 4.24a and 4.24b show the $\sigma_{T_{PERIOD}}$ as a function of RSD_{SIF} for the measured ADC data. In the optimal combination of the sampling speed (4 GS/s), signal frequency (1 GHz), and SIF/RSD (5.75%/4.34%), a $\sigma_{T_{PERIOD}}$ of approximately 454 fs has been measured. It has to be noted that $f_{SIG} = 1 \text{ GHz}$ violates the postulated BW_{limit} introduced with Equation 4.7 for pulse waveforms. Signals with a sine-like waveforms have non-linear smooth edges and thus do not suffer the abrupt changes in the signal slope like the pulse-type signals. However, considering the BW_{limit} for $f_{SAMP} = 4 \text{ GS/s}$, the lowest measured $\sigma_{T_{PERIOD}}$ was 505 fs at $f_{SIG} = 800 \text{ MHz}$.

While the RSD might be good for visualization of the data in terms of the SIF parameter, it offers very little prediction. In fact, we can see from Figures 4.24a and 4.24b that there is no trend line in terms of signal frequency. However, if we plot $\sigma_{T_{PERIOD}}$ as a function of the f_{SAMP}/f_{SIG} ratio as shown in Figure 4.25, the signal period resolution is very good at integer values of this ratio. In addition, the peaks of the $\sigma_{T_{PERIOD}}$ worsen considerably towards higher values of the signal frequency, which is a consequence of the non-linear edge of the sampled signal.

While the RSD might be good for data visualization in terms of the SIF parameter, it offers very little prediction in terms of expected timing resolution as a function of the signal frequency. In fact, we can see from Figures 4.24a and 4.24b that there is no trend line in terms of signal frequency. However, if we plot $\sigma_{T_{PERIOD}}$ as a function of the f_{SAMP}/f_{SIG} ratio as shown in Figure 4.25, the

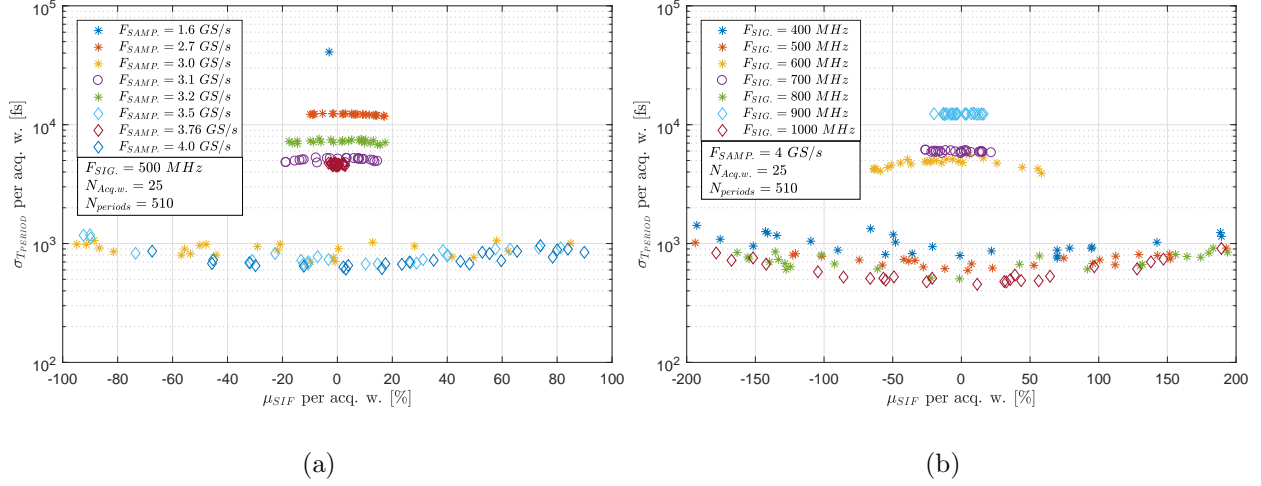


Figure 4.23: (a) Extracted signal period resolution $\sigma_{T_{PERIOD}}$ from the ADC data as a function of μ_{SIF} for available ADC sampling speeds at a signal frequency of 500 MHz. (b) Extracted signal period resolution $\sigma_{T_{PERIOD}}$ from the ADC data as a function of μ_{SIF} for signal frequencies ranging from 400 MHz to 1 GHz at the ADC sampling speed of 4 GS/s.

signal-period resolution is very good at integer values of this ratio. In addition, the peaks of the $\sigma_{T_{PERIOD}}$ worsen considerably towards higher values of the signal frequency, which is a consequence of the non-linear edge of the sampled signal.

We can thus conclude that in the ADC measurements, the extracted period resolution is dominated by the apparent rotation of the phase due to the fractional relationship of the sampling speed

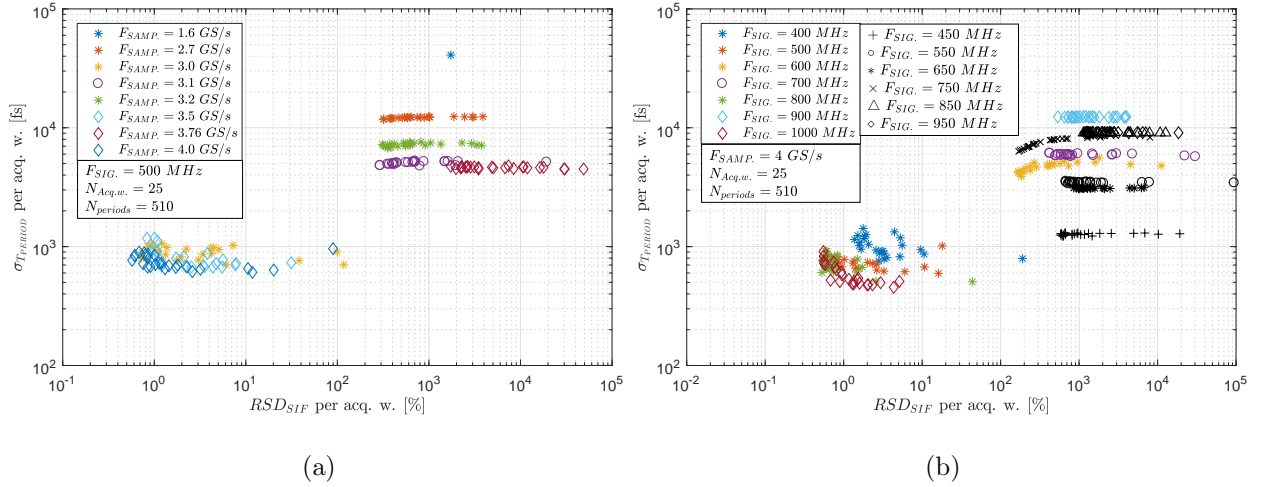


Figure 4.24: (a) Extracted signal period resolution from the ADC data as a function of RSD_{SIF} for the available ADC sampling speeds at the signal frequency of 500 MHz. (b) Measured (colored points) and simulated (black points) signal period resolution as a function of RSD_{SIF} for signal frequencies ranging from 400 MHz to 1 GHz at an ADC sampling speed of 4 GS/s.

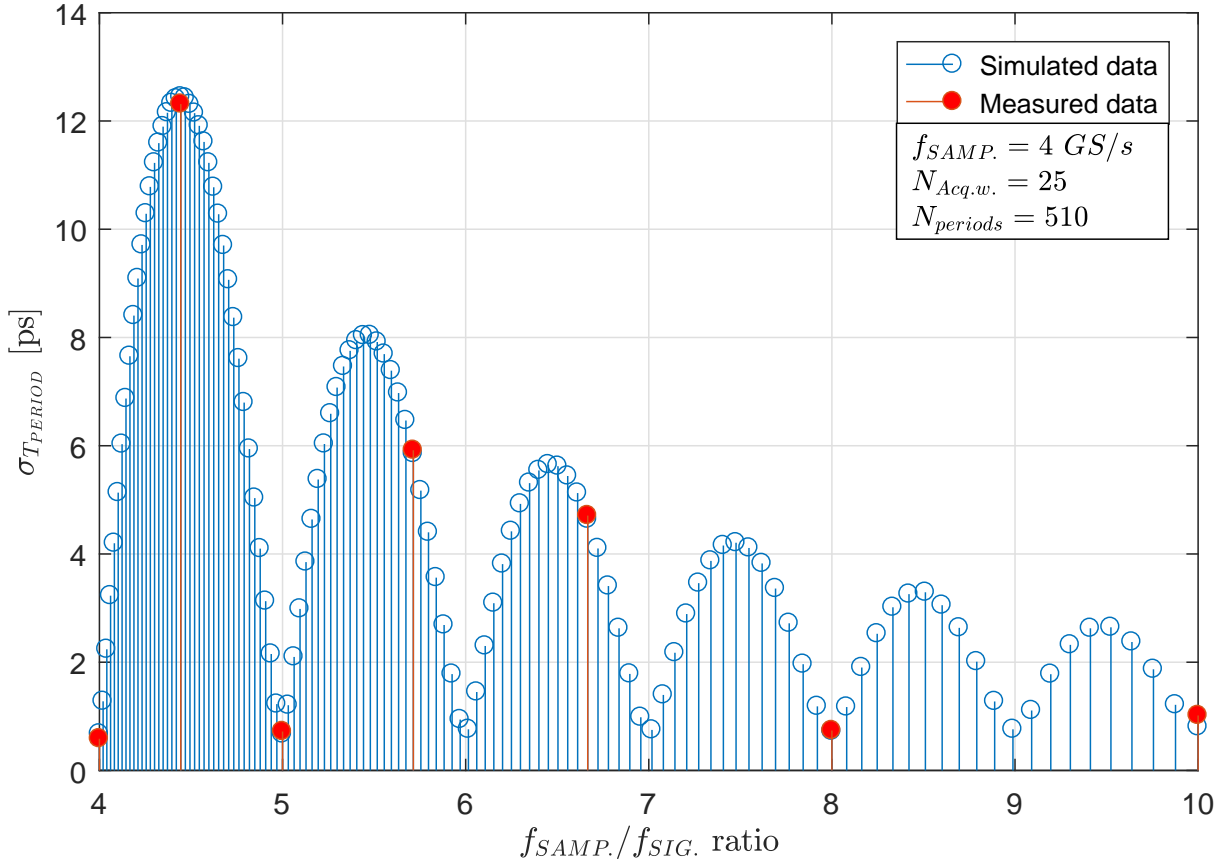


Figure 4.25: Measured and simulated signal period resolution as a function of the $f_{SAMP.}/f_{SIG.}$ ratio for signal frequencies ranging from 400 MHz to 1 GHz at an ADC sampling speed of 4 GS/s.

and signal frequency as well as the non-linearity of the signal edge.

ADC Measurement Results (Synchronous Mode)

Creating a perfectly synchronous signal with the sampling time base was achieved by modifying the ADC evaluation board by splitting the sampling clock. One part was driving the ADC itself, while the other was taken off-board to a custom PCB board, which was designed to divide the sampling frequency by eight, filter, and amplify the signal. This signal was then measured by the ADC. Figure 4.26 shows the used setup. Phase changes were introduced by different lengths of the cable connecting to the ADC. Figure 4.27 shows the results of the measurements along with the corresponding SWG data. Very good agreement is observed.

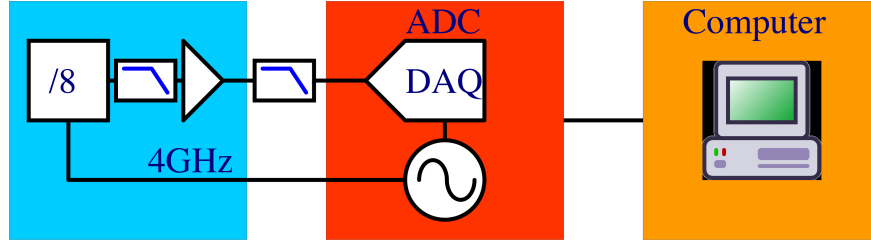


Figure 4.26: Test bench setup block diagram (synchronous mode).

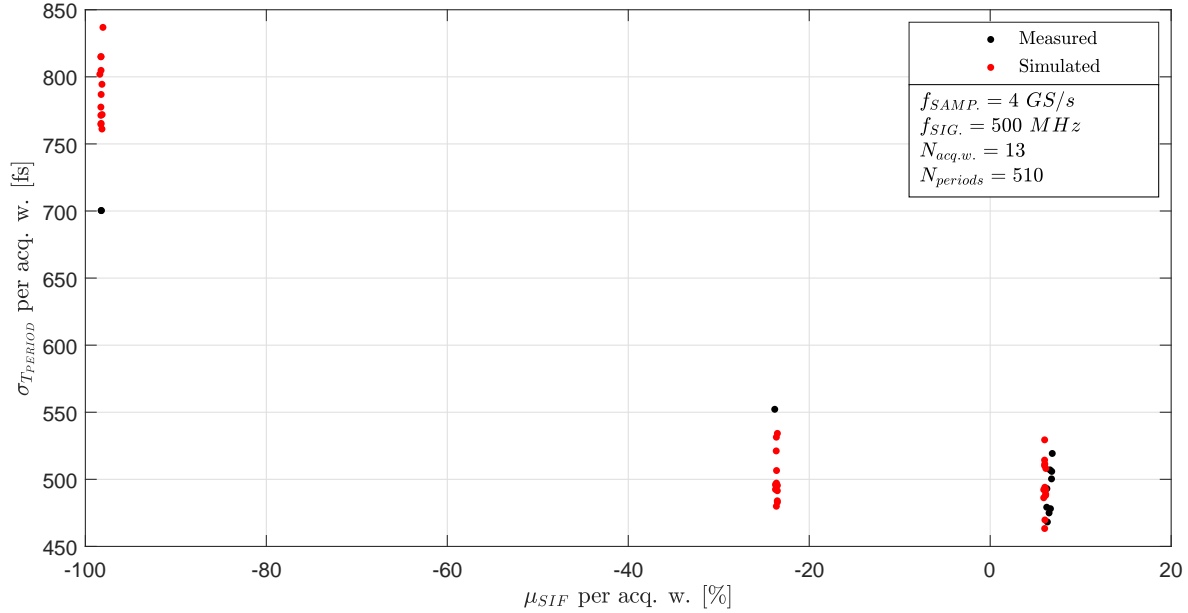


Figure 4.27: Measured and simulated $\sigma_{T_{PERIOD}}$ as a function of μ_{SIF} for the synchronous mode of operation.

As expected, working in the synchronous mode and having the μ_{SIF} very close to zero gives the best results. In fact, the best measured $\sigma_{T_{PERIOD}}$ was 468 fs.

4.6 Alternative Waveform Shapes

The noise in the system is dominated by the thermal noise, which has a normal distribution in time domain and a uniform distribution over the frequency spectrum. Thus, a narrower frequency spectrum in terms of the analog bandwidth would mean lower integrated noise. Gaussian pulses have the spectrum components all the way to the DC. However, the second derivative of the Gaussian pulse, also known as the Mexican hat or the Ricker wavelet, does not [70]. Figure 4.28a shows a time domain representation of a Ricker wavelet with the needed rise time (t_R) of 120 ps. Figure 4.28a shows the frequency spectrum of this particular waveform.

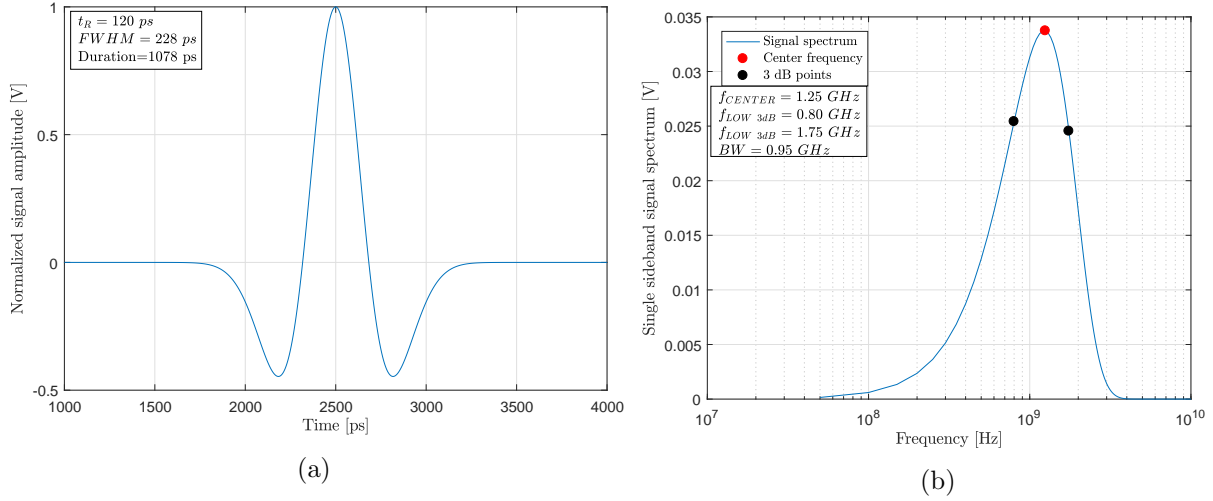


Figure 4.28: (a) Time domain plot of a Ricker wavelet with a rise time of 120 ps and an approximate duration of 1.1 ns. (b) Single sideband spectrum of the Ricker wavelet with the maxima and 3 dB points displayed.

The frequency span of this pulse ranges between 0.8 GHz and 1.75 GHz, thus requiring an analog bandwidth of approximately 950 MHz. The integrated thermal noise floor in this bandwidth is approximately 39 μV @ $Z = 50\ \Omega, T = 300\ K$, while the integrated thermal noise floor for $BW = 3\ GHz$ is approximately 69 μV @ $Z = 50\ \Omega, T = 300\ K$. Thus the noise is reduced by a factor of ≈ 1.8 . Furthermore, the analog bandwidth requirement is considerably loosened.

However, with a pulse duration of approximately 1.1 ns the current transmission line does not have sufficient length to hold two such pulses as it is required for the TVD specification.

4.7 Summary

The results of this study clearly show that noise, jitter, and distortion are the dominant sources of error, which was somewhat expected. However, the coupling mechanisms by which these sources of error impact the timing resolution on the femtosecond level were not obvious and/or quantified before. In light of this, crucial knowledge and insight has been obtained for the development of fast, wideband RF systems that can achieve femtosecond timing resolution. Some of the gained insights are summarized below:

- Limited vertical resolution affects the sampled data in the form of distortion. Its effect gets masked by the noise present in the systems at higher numbers of bits. In our case, increasing the number of bits beyond ten does not lead to significant improvement.
- Jitter couples directly into the sampled data. The most effective way of decreasing its effect is by implementing a very low jitter clock buffering and distribution circuit (added jitter no more than 40 fs) as well as by using a clean clock source (source jitter no more than 20 fs).
- The amplitude noise of the system is a greater contributor to the over-all timing resolution than the sampling speed and/or analog bandwidth. A reduction in the bandwidth would lead to lower integrated noise. New signal shapes with narrower frequency spectra, such as the Richer wavelet (Mexican hat wavelet) are being investigated.
- The ability to adjust the sampling speed according to the known input signal bandwidth and the signal frequency can greatly improve the timing resolution, provided that we have:
 - advanced knowledge of the signal waveform,
 - well controlled signal sources.
- If synchronous operation is possible, a favorable and fixed phase offset can be achieved in parallel with a favorable f_{SAMP}/BW ratio, thus providing the optimal operating point for the measurements.
- The choice of the extraction algorithm greatly affects the timing resolution. An algorithm based on a linear fit will produce good results only on a linear edge, which is hard to achieve in practice. Therefore, a curve-fitting algorithm can considerably improve the extracted timing resolution. However, this comes at the expense of greater processing power and higher data rates.

Furthermore, good agreement between the simulated and the measured results supports the validity of the SWG. This simulation tool thus provides a platform to test different algorithms

coupled with various hardware settings where limiting cases and underlying mechanisms can be easily explored without some of the complications of a real test bench. In addition, real-life tests can be limited to setups that show promising results in simulations, thus considerably lowering the R&D costs.

From the presented analysis of the simulation results, we can conclude that the estimations provided by Equation 2.8 are not sufficiently reliable at the femtosecond timing resolution level. Therefore, using the SWG has been essential in estimating the hardware response and determining the RFpix specifications.

CHAPTER 5

RFPIX PROTOTYPE DESIGN

The first prototype of the TVD readout ASIC is called the RFpix1. This chapter is centered on the development of RFpix1 analog front-end. The architecture and the specifications are presented first and followed by an in-depth discussion on individual sub-circuits that make up the RFpix1 analog front-end. All presented simulation results include layout parasitics.

5.1 RFpix1 Specifications and Architecture

The RFpix1 prototype is being designed specifically to reach the primary goal of the target extracted timing resolution of 100 fs or less. All of the other requirements like the form factor, data rate and power consumption have been reduced in priority. However, it is worth noting that even though these requirements do not represent the main drive for the design of the prototype, they were still taken into account where possible.

The RFpix1 differs from the final detector version, as listed below:

1. The RFpix1 has only eight out of 126 channels;
2. The readout logic is simplified to the point of not containing the sampling discrimination logic used to reduce the data throughput;
3. The die is to be wire bonded and enclosed in a regular surface mount leaded flat package, while the detector version is to be bonded directly to the VDS substrate using flip-chip technology;
4. Due to the packaging, the input channel is buffered to achieve the desired analog bandwidth. Consequently, the prototype power consumption is expected to be higher than the final version.

Table 5.1 summarizes the RFpix1 design specifications, while Figure 5.1 shows the simplified functional block schematic of the RFpix1.

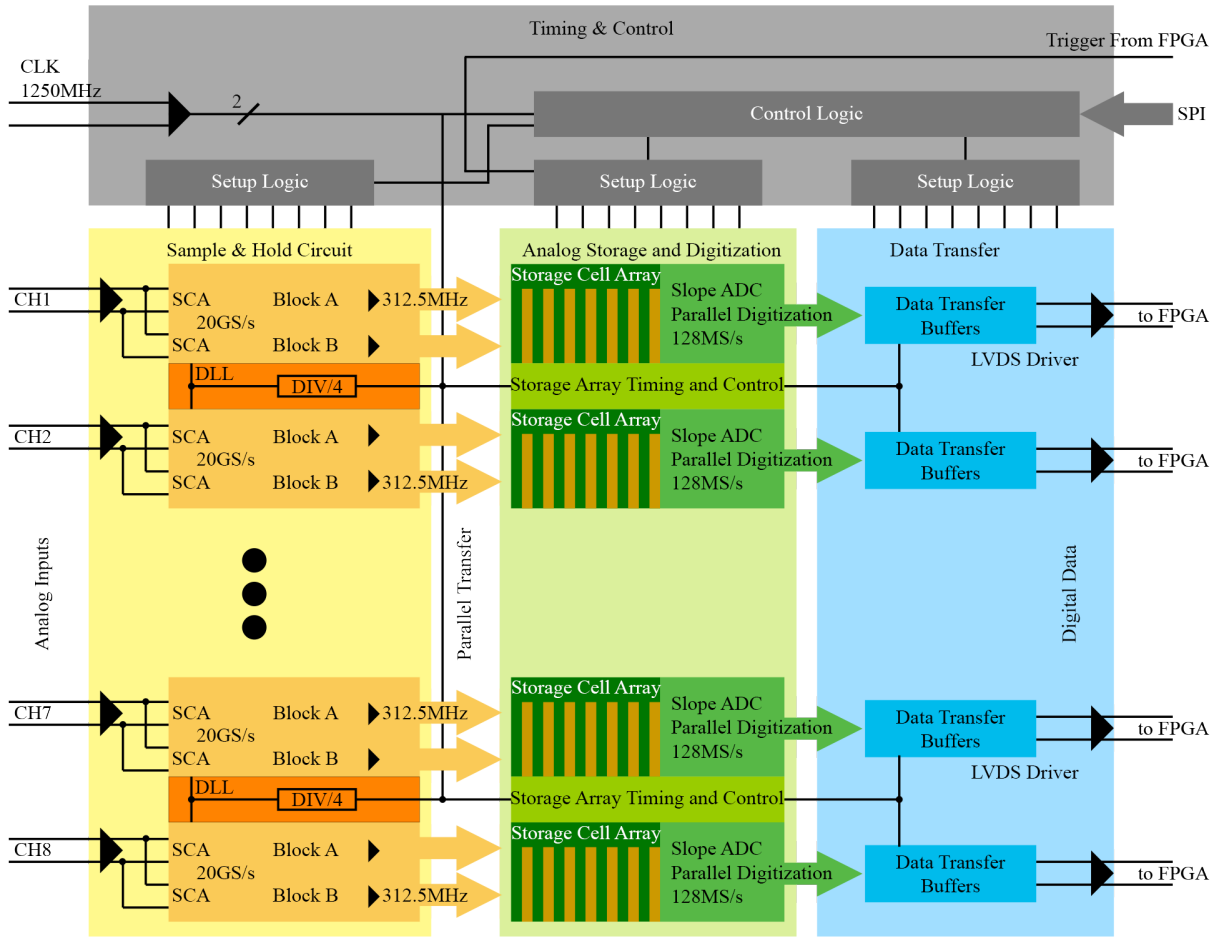


Figure 5.1: Simplified functional block schematic of the RFpix1.

Table 5.1: Baseline RFpix1 design specifications.

<i>Parameter</i>	<i>Value</i>
Sampling period	$50 \text{ ps @ } 20 \text{ GS/s}$
Analog bandwidth	$\approx 3 \text{ GHz}$
Input referred noise	$\leq 0.5 \text{ mV}_{RMS}$
Added jitter per channel	$\approx 40 \text{ fs}$
Number of bits	12
ENOB	≥ 10
Number of channels	8
Power consumption per channel	TBD
Data rate per chip	285 Mbits/s or 8.88 Mbits/s per channel

The prototype is being designed in a 130 nm CMOS technology node. The three main reasons for this choice are:

1. Very good performance in terms of radiation hardness has been proven [47] for this technology node;
2. Best tradoff between power consumption, jitter, size, and cost [71];
3. Technology nodes with smaller feature sizes experience a shrink in the dynamic range below the minimum required range (1 V_{pp}).

As mentioned in chapter two, the RFpix architecture can be subdivided into four sections:

- Sample & Hold;
- Analog Storage and Digitization;
- Data Transfer;
- Timing & Control.

A channel is composed of two blocks with 64 sampling cells each. That is, 128 sampling cells in total. The SCA cell follows a differential topology and it is specifically designed for low distortion and high tracking bandwidth. The differential configuration of the SCA also helps in terms of crosstalk mitigation and reduction of noise coupling. At the same time, it turns the amplitude dependent voltage error due to charge injection into a virtual gain of the sampling cell. The sampling cell is composed of a differential switch followed by a sampling capacitor and a buffer.

The individual sampling cell switches represent a capacitive load to the input transmission line. This load in conjunction with the series inductance of the wire bonds forms an LC filter circuit. To maximize the input analog bandwidth, the load capacitance needs to be minimized, which entails smaller transistors. However, smaller transistors have higher track-mode resistance, thus lowering the sampling cell tracking bandwidth. To separate this two mutually exclusive requirements, every RFpix1 channel has an input unity gain buffer that isolates the input pad from the sampling array. This way higher capacitive loads become sustainable, thus providing a degree of freedom for the sampling cell design.

The sampling array is driven by strobe signals that are generated by a fully differential two-level delay-locked loop (TLDLL). A multilevel topology has been chosen in order to decrease the added jitter of the TLDLL. The TLDLL outputs 64 strobe signals along with a block select signal. The block select signal ensures that only one of the SCA blocks is sampling at the time, while the other one is transferring the sampled voltage to an analog storage array.

The analog storage array follows a single-ended topology and it is composed of 2048 (32×64) storage cells. Each SCA block is linked to its own storage array block for a total buffer depth of 32 storage cells per channel. The storage array is followed by a single-slope or Wilkinson ADC.

The single-slope ADC is designed as a distributed circuit. The comparators are integrated into the storage array, while the ramp for comparator triggering is sourced from a single ramp generator, which is composed of an input/output rail-to-rail operational amplifier (op-amp) configured as an integrator. That is, a DAC is used to create a voltage difference between the op-amp inputs. This voltage difference is then integrated in time creating a linear ramp. The speed and the slope polarity depend on the magnitude and polarity of the voltage difference, making the ramp is highly adjustable from $0.5 \mu s$ to $2 \mu s$. The digitizing logic triggers the ramp generation at the same time as it triggers a 12-bit counter. The counter outputs are fed into 64 registers (one for each storage cell within the window). These registers are latched by the storage cell comparators, thus retaining the counter value and producing the 12-bit word representing the sampled values. The maximum speed of each distributed ADC part is $2 MS/s$. Since they running in parallel, the maximum ADC speed of each channel is $128 MS/s$. Every channel has a dedicated ADC, thus making them independent from each other.

At the end of each digitization cycle, the contents of the ADC latches are shifted to an output buffer and is transmitted off-chip by a dedicated LVDS driver. There are two LVDS drivers per channel. One transmits the clock, while the other transmits the data. With a 12-bit ADC, buffer depth of $5 \mu s$, and the system trigger of $30 kHz$, the overall data throughput is estimated to be around $2 Mbits/s/channel$. The data is expected to be received by and external FPGA.

The prototype chip is to be controlled by a series of shift registers and digital-to-analog converters, which form the control portion of the Timing & Control section of the ASIC. This circuits are expected to be programmed by an FPGA through a serial peripheral interface (SPI).

In the following sections, we will be focusing on the RFpix1 analog front-end. More specifically, the switched capacitor array, the delay-locked loop, the storage array, and the corresponding sub-circuits.

5.2 Switched Capacitor Array

Each SCA cell is composed of two sampling cells denoted as A and B, thus forming the two sampling blocks and a differential clock driver that ensures proper alignment of the sampling strobes and proper switching between the two sampling blocks.

To improve upon the results presented in chapter three and meet the RFpix1 requirements, a new sampling cell architecture is required. The chosen differential topology decreases the effective coupling of the jitter from the TLDLL, provides crosstalk immunity (decreased common-mode noise coupling), and decreases the required single-ended input dynamic range from $1 V_{PP}$ to $0.5 V_{PP}$.

5.2.1 Sampling Cell, First Stage (Sampling Switches and Sampling Capacitor)

The sampling cell architecture is shown in Figure 5.2a. The structure follows a differential topology with the two switches denoted as P and N respectively. These switches are composed of two complementary CMOS transistors. The switches provide an electrical pathway from the transmission line to the sampling capacitor and the sampling cell buffer.

The overall track-mode switch resistance along with the on-state resistances of the individual transistors and the associated tracking bandwidth are shown in Figure 5.2b.

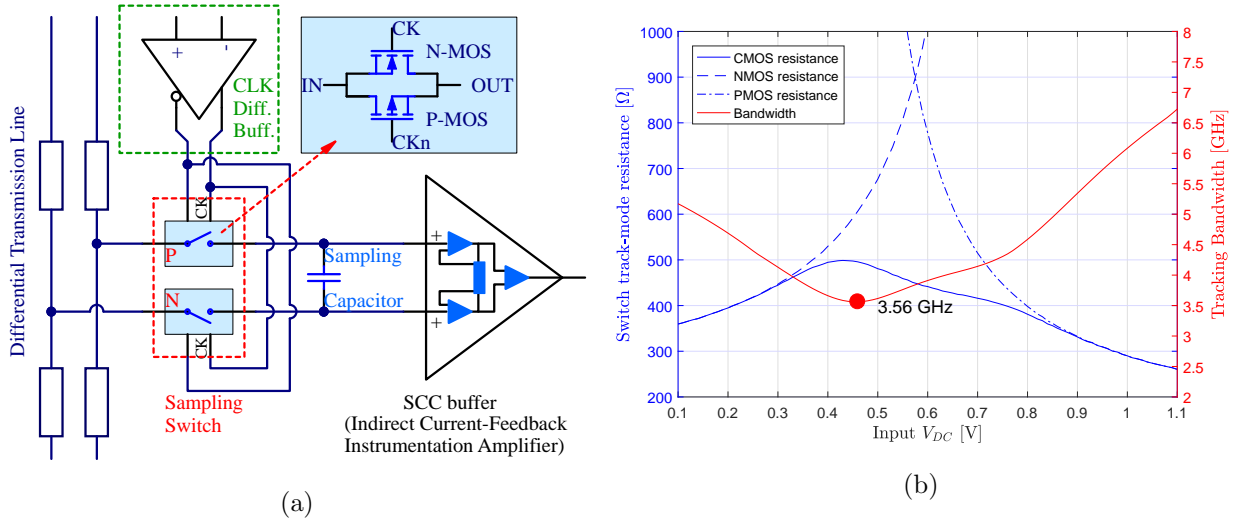


Figure 5.2: (a) Simplified RFpix1 sampling cell schematic showing the differential input structure with the input switch, sampling capacitor, the differential clock buffer, and the sampling cell buffer, which is an indirect current-feedback instrumentation amplifier. (b) RFpix1 sampling cell track-mode switch resistance and tracking bandwidth as functions of input V_{DC} .

The direct comparison between the PSEC4 and RFpix1 sampling cells in terms of track mode resistance and tracking bandwidth is shown in Figures 5.3a and 5.3b respectively.

Using low-threshold voltage transistor for implementing the RFpix1 switch, the track-mode

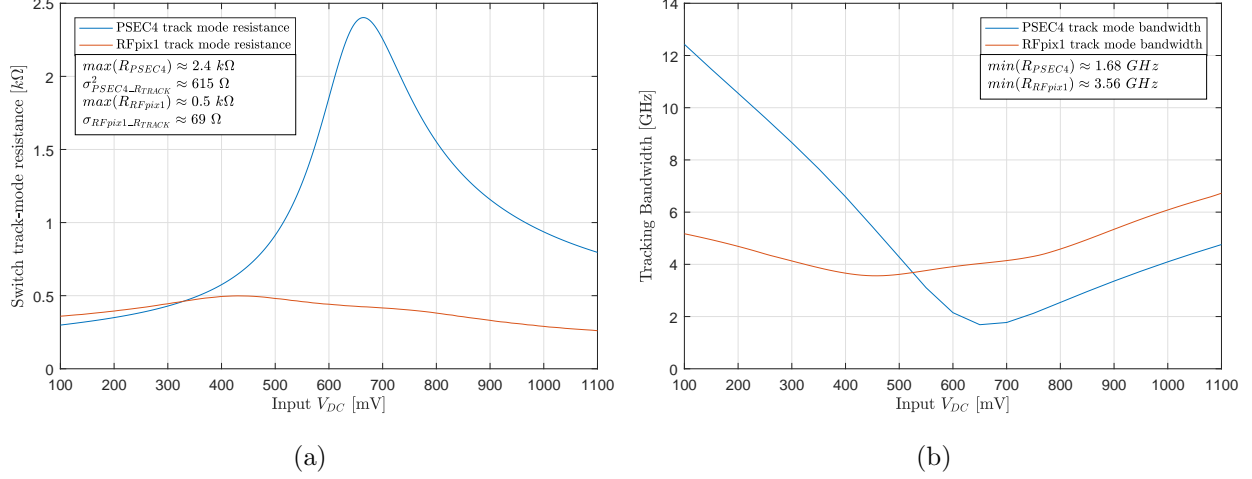


Figure 5.3: (a) Comparison between the RFpix1 and the PSEC4 sampling cell track-mode switch resistances as functions of input V_{DC} . (b) Comparison between the RFpix1 and the PSEC4 sampling cell tracking bandwidths as functions of input V_{DC} .

resistance variance has been lowered considerably ($\sigma \approx 69 \Omega$) in comparison with PSEC4 switch track-mode resistance variance ($\sigma \approx 615 \Omega$). In addition, the lowest tracking bandwidth is approximately 3.56 GHz , leading to a significant improvement of the signal response. Figure 5.4a shows the THD at $V_{DC} \approx 600 \text{ mV}$ for input signal frequencies of 1.5 GHz , 2 GHz , and 3 GHz respectively. The differential topology of the SCC provides with a significant reduction in distortion.

Simulations of the load capacitance towards the transmission line are shown in Figure 5.4b. The load capacitance of the sampling cell in hold mode is approximately 11 fF , while the total load capacitance seen at the input of the sampling cell in track mode varies from as low as 47 fF to as high as 57 fF . The extrapolated sampling cell capacitance is around 34 fF .

The noise of the sampling cell needs to be considered for the track and hold mode separately:

1. **Track mode noise:** In track mode, the noise is dominated by the thermal noise of the on-state switch resistance integrated over the tracking bandwidth. The sampling cell output referred single sideband noise spectrum is shown in 5.5a. The in-band single-ended voltage noise density (e_n) is approximately $2.5 \text{ nV}/\sqrt{\text{Hz}}$. Consequently, the integrated noise floor is approximately $232 \mu\text{V}$.
2. **Hold mode noise:** In hold mode, the noise is dominated by the thermal excitation of the charge electrons on the sampling capacitor also called the kTC noise and is independent of the bandwidth [48]. The kTC noise is calculated by using Equation 5.1,

$$\Delta u = \sqrt{\frac{k_B T}{C}} \quad (5.1)$$

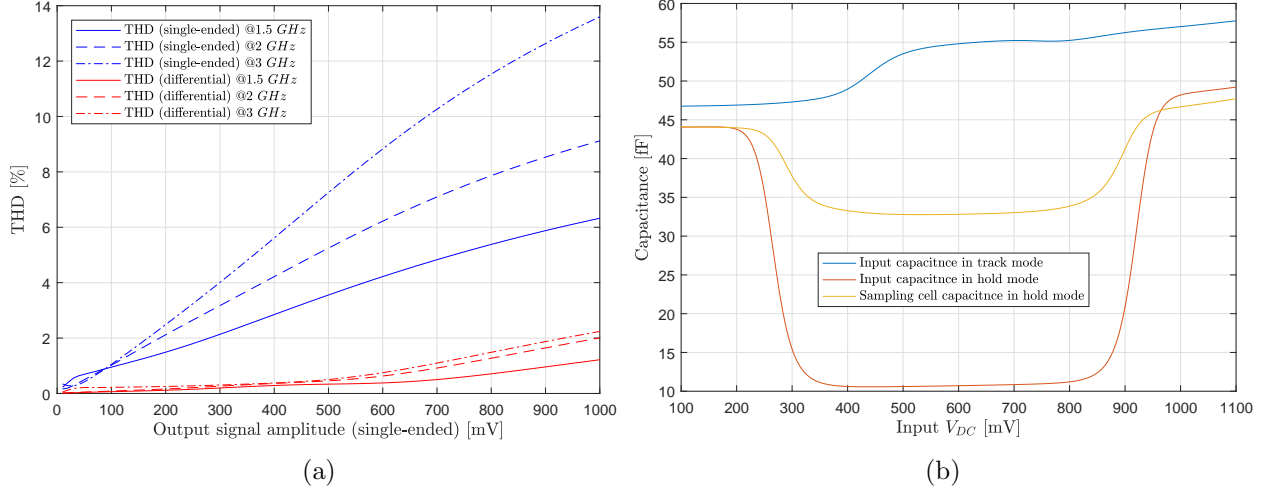


Figure 5.4: (a) RFpix1 sampling cell THD curves as functions of signal amplitude at different input signal frequencies. (b) Input and sampling cell capacitances in track and hold mode as functions of input DC voltage.

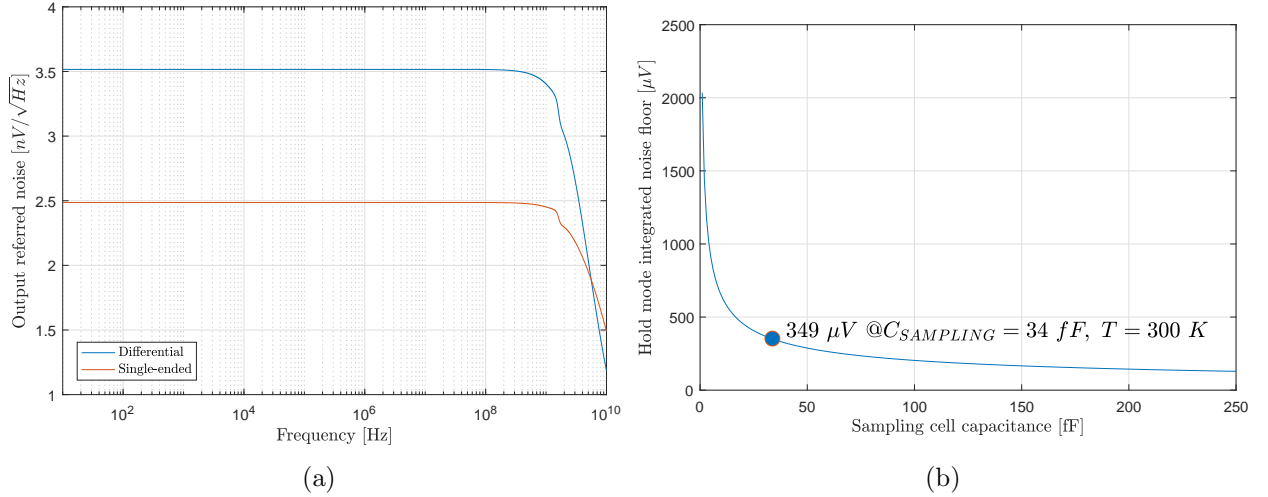


Figure 5.5: (a) Track mode single sideband (SSB) noise spectrum of the sampling cell. (b) Sampling cell integrated noise floor as a function of sampling capacitance.

where k_B is the Boltzmann constant and T is the temperature in Kelvin. Figure 5.5a shows the noise floor of the sampling cell in hold mode a function of capacitance. In our case, considering a sampling cell capacitance of 34 fF , the kTC noise of the sampling cell hold-mode is approximately $349 \mu V_{RMS}$.

5.2.2 Sampling Cell, Second Stage (Sampling Cell Buffer)

When both switches close simultaneously, the sampling capacitor retains the differential voltage between its two terminals. In order to transfer this voltage without disturbing the charge and consequently the voltage on the sampling capacitor, a buffer with high impedance inputs is needed. Furthermore, the buffer amplifier has to convert the sampling cell differential voltage into a proportional single-ended voltage that can be stored in the analog storage array.

Instrumentation amplifiers provide high impedance inputs and at the same time offer differential to single-ended conversion. A typical instrumentation amplifier configuration is the classic three op-amp configuration as shown in Figure 5.6. The two input op-amps (on the left) are configured in a non-inverting configuration, while the output op-amp is in a difference configuration (on the right). Op-amps do not have accurate open-loop gains. For this reason, the three op-amps instrumentation amplifier needs feedback resistors. The gain of the circuit is defined by the bridge resistor (R_{gain}). These feedback resistors are crossing the current-source isolation barrier and thus

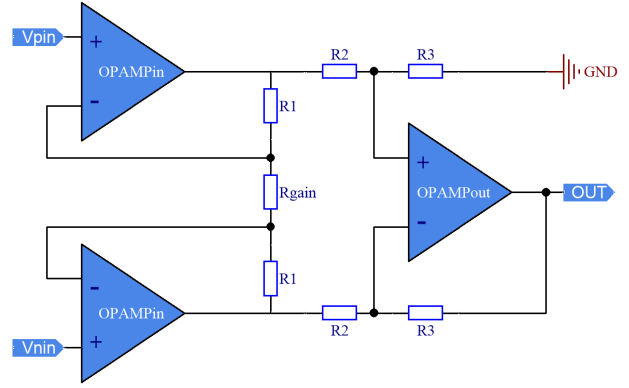


Figure 5.6: Instrumentation amplifier realization with three op-amps.

lowering the value of the common mode rejection ratio (CMRR). In addition, the two input amplifiers see two different loads when driving the output op-amp. This leads to asymmetries in the overall amplifier response. Compensating these asymmetries entails the usage of either strong output drivers or high resistance in the feedback. Strong output drivers lead to higher power consumption while high resistor values lead to higher thermal noise. To satisfy the low power and low noise requirements of the sampling buffer a different instrumentation amplifier topology has to be adopted. In our case, an indirect current-feedback topology has been adopted and modified to adhere to the sampling buffer requirements [72].

Indirect Current-Feedback Instrumentation Amplifier

A standard indirect current-feedback instrumentation amplifier topology is shown in Figure 5.7. Such a topology was first used to implement a method for frequency compensation and later to implement an instrumentation amplifier. The principle of operation is based on two groups of equal currents, where I_1 to I_4 form one group and currents I_5 and I_6 form the other group. The loop amplifier A_1 forces the differential collector current from Q_1 and Q_2 into Q_3 and Q_4 by driving the required voltage V_{fb} at the inputs of Q_3 and Q_4 . Another way of looking at this circuit is that the feedback transconductance amplifier, implemented with Q_3 , Q_4 , and R_2 , together with the loop amplifier A_1 , form a two-stage operational amplifier. A replica transconductance amplifier implemented with Q_1 , Q_2 , and R_1 is added to this operational amplifier and is used as a new input stage to implement the indirect current-feedback [7].

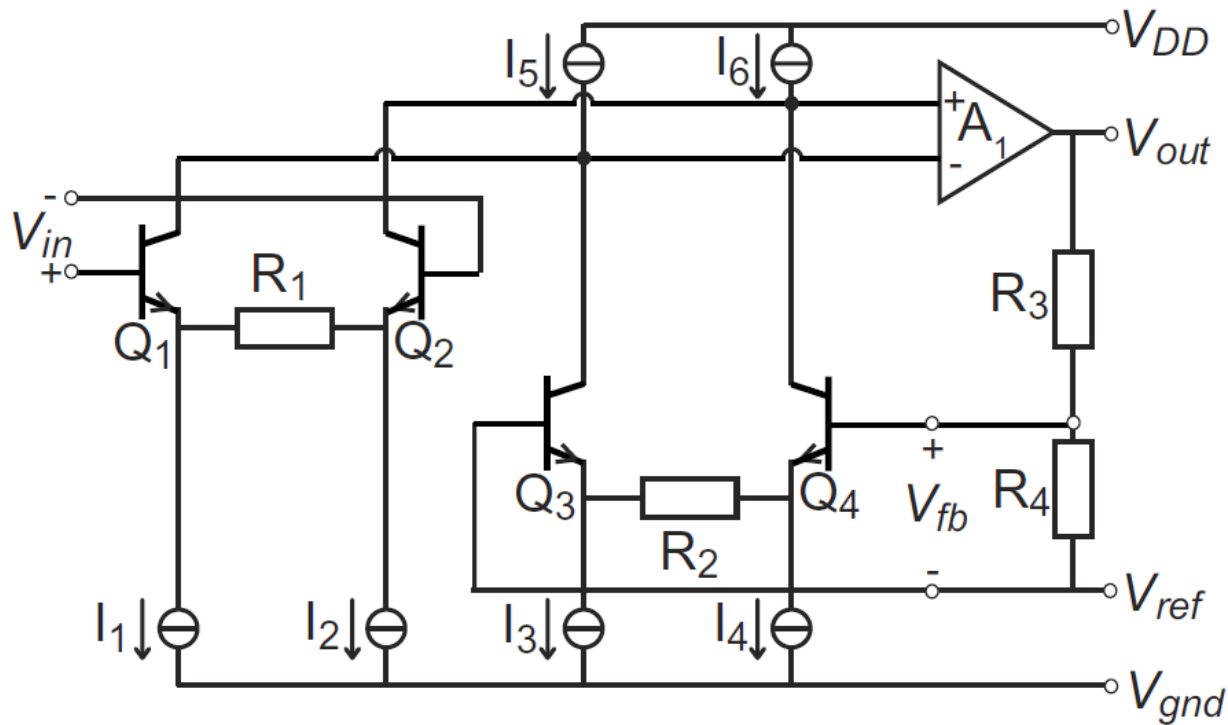


Figure 5.7: Schematic showing an indirect current-feedback instrumentation amplifier. The amplifier is implemented using bipolar transistors, but the same principle applies to a CMOS variant [7].

Sampling Cell Buffer

The sampling cell buffer requires a flat gain response over an output dynamic range of $1 V_{PP}$. The differential to single-ended conversion provides with an effective signal gain of $2 V/V$. At the same time, the supply rail in this technology is $1.2 V$, thus making the output dynamic range almost rail-

to-rail. Figure 5.8 shows the transistor level schematic of the sampling cell buffer. The principle of operation of the shown topology is still that of an indirect current-feedback instrumentation amplifier, however, the transistor level realization is somewhat different:

- The differential pairs are designed with a single current source, which leads to lower current consumption.
- The left side of the circuit is composed of two sections: the input and the feedback sections. Both sections are composed of complementary N-type and P-type differential pairs respectively. However, only the feedback section requires the implementation of the complementary differential pairs. That is, the feedback section is directly driven by the output, thus needing to operate over the full output dynamic range of $1 V_{PP}$. Conversely, the input section is driven differentially and only needs to operate over a voltage input dynamic range of $0.5 V_{PP}$. However, to maintain symmetry and thus linearity the input section has also been realized with complementary differential pairs.
- The middle section of the circuit is the cascode stage. The cascode stage has several functions: merging the currents from the N-type and the P-type differential pairs from both the input and feedback sections, increasing the amplifier open-loop gain, providing the necessary voltage offset between the output transistors in order to drive them in the AB regime.
- The output stage is composed of an inverter circuit, which is driven in the AB regime. The benefit of this configuration is the ability to provide high current drive during transients, while maintaining a low quiescent current in steady-state. The drawback is the large variation of the transconductance of the output dynamic range, thus compromising the amplifier stability.
- To stabilize the overall transient response of the amplifier a feedback capacitance has been introduced in the form of a NMOS transistor. The transistor gate is connected to the output, while the source and drain are connected to the drive signal of M22.

The DC response of the sampling cell buffer is shown in Figure 5.9a. The linearity over the dynamic range varies considerably. However, the area of most interest in terms of timing extraction is at the center, where the deviation is around $\pm 5 mV$. Figure 5.9b shows the transient response of the amplifier. Due to the AB class operation, the transient response experiences a fast rising edge and then a slow settling time. In order to achieve reasonable settling times over most of the dynamic range, high small signal bandwidth is needed. Figure 5.10a shows the small signal bandwidth of the sampling cell buffer over the dynamic range. The bandwidth is the highest ($\approx 3.9 GHz$) at the center of the dynamic range, where both the N-type and P-type differential pairs have the highest transconductances. In addition, the closed loop gain shows a resonance peak, which contributes to the high bandwidth. The closed loop gain and phase responses at the input DC voltage of $600 mV$

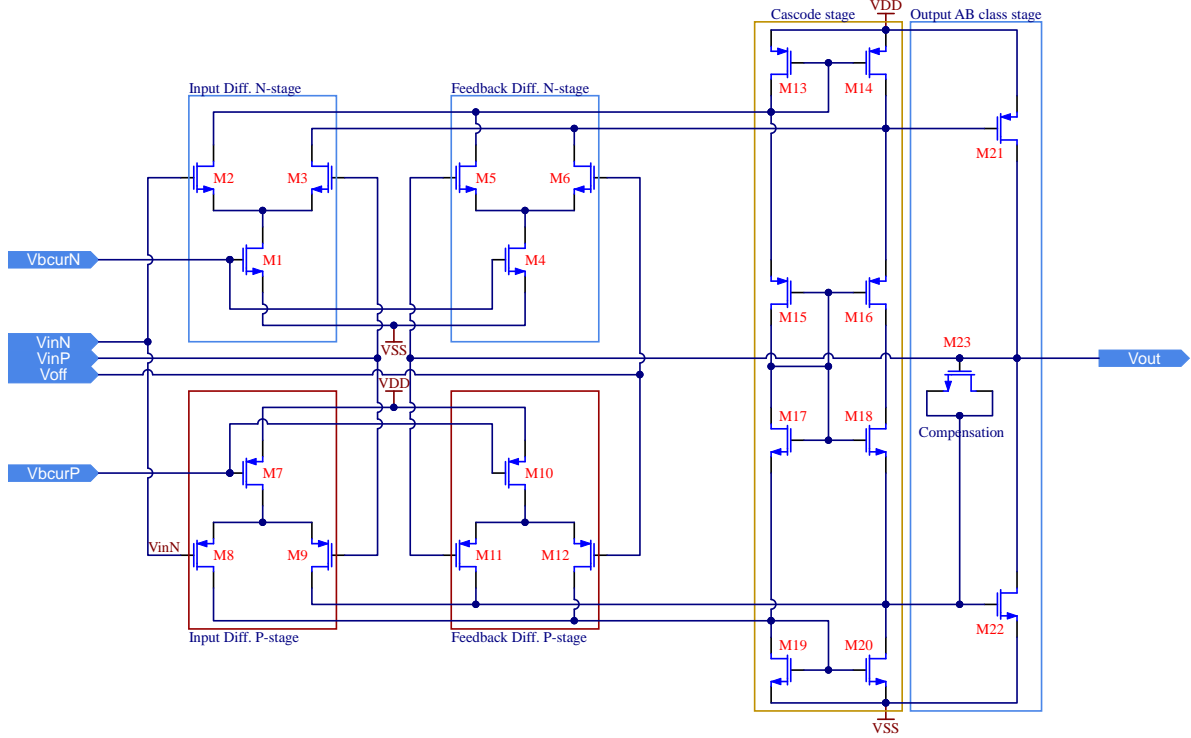


Figure 5.8: Transistor level schematic of the sampling cell buffer.

are shown in Figure 5.10b. The phase margin is very small, which explains the small ringing seen in the transient response. However, upon inspection of the poles of the transfer function for two input DC offsets, we notice that all of the poles lay in the left half-plane of the pole diagram, thus confirming the stability of the amplifier. The poles at two different DC offsets are shown in Figure 5.11a. The large imaginary component of the conjugate pair close to the imaginary axis explains the ringing in the transient response at the center of the dynamic range. This non-linearities give rise to distortion.

The total harmonic distortion as a function of input signal amplitude is shown in Figure 5.11b. The noise spectrum of the amplifier is shown in Figure 5.12a. The $1/f$ cutoff is around 10 kHz , while the noise floor is dominated by white noise with an rms value of approximately $8.7\text{ nV}/\sqrt{Hz}$. Due to the large bandwidth, the integrated noise floor is approximately 1 mV_{RMS} . This noise level is somewhat high for our application. However, the common mode rejection ratio is good enough to diminish the impact of the signal noise, thus providing some room for error. That is, the signal scales with a gain of two while the noise scales with a gain of $\sqrt{2}$. The CMRR is shown in Figure 5.12b.

The overall power consumption is difficult to estimate because it varies with the event rate due to significant differences between the quiescent current draw and the transient current draw. Figures 5.13a and 5.13b respectively show the quiescent current draw as a function of input DC

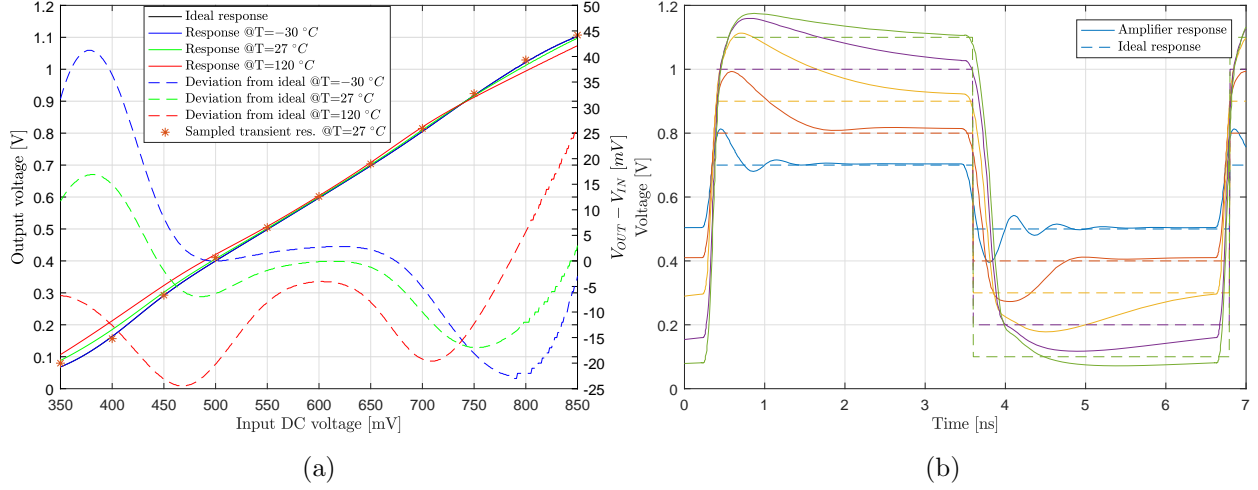


Figure 5.9: (a) Sampling cell buffer DC response at three different temperatures. (b) Sampling cell buffer transient response.

voltage and the transient current for three different output amplitude swings. The average current draw can be estimated by considering that the average channel occupancy is 0.57 %, thus the average time between hits is 351 ns. Consequently, most of the time the amplifier has a zero input voltage and the average current draw is approximately 272 μA . Considering 128 such buffers in a single channel, the average current draw of this buffers amounts to approximately 34.8 mA per channel. The power added efficiency during the maximum output voltage swing is approximately 77.8 %.

The layout of the sampling cell buffer is shown in Figure 5.14, with a size of 17.395 $\mu m \times$ 14.85 μm .

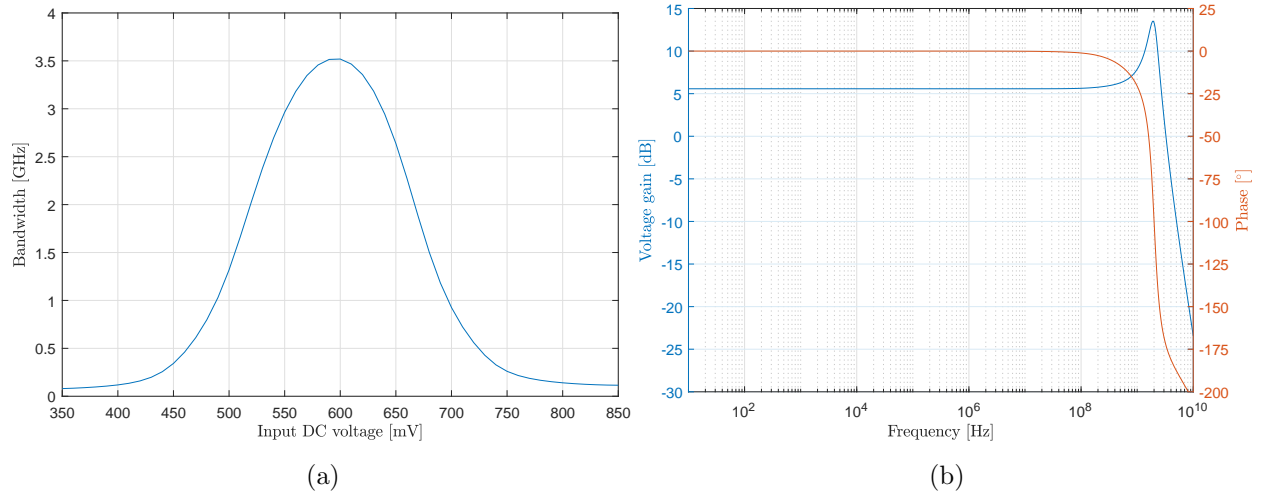


Figure 5.10: (a) Sampling cell buffer small signal bandwidth as a function of input DC voltage. (b) Sampling cell buffer gain and phase at an input DC voltage of 600 mV.

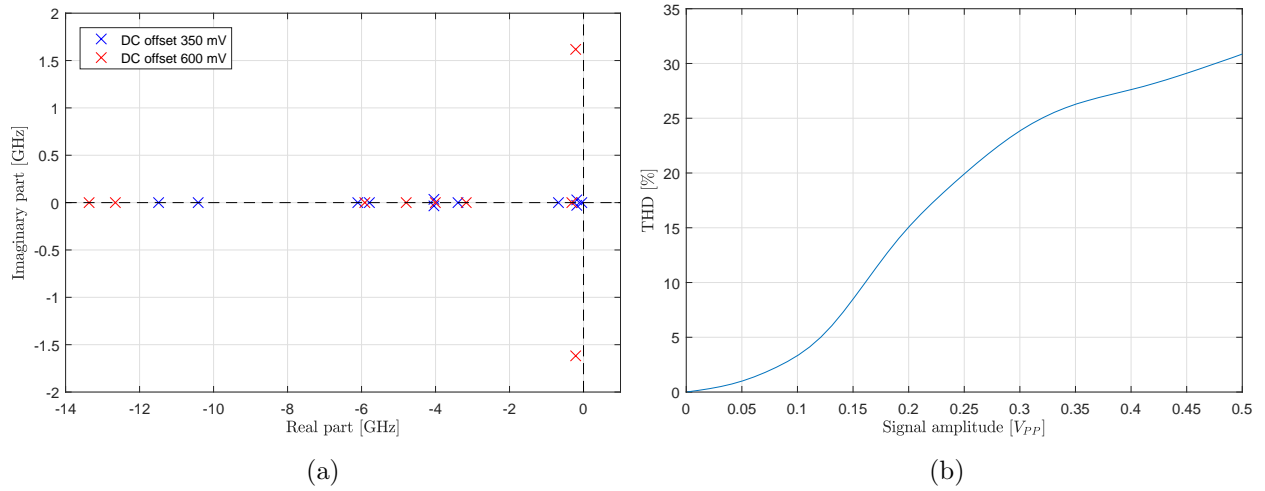


Figure 5.11: (a) Pole diagram of the sampling cell buffer for two input DC offsets. (b) Sampling cell buffer total harmonic distortion as a function of input signal amplitude.

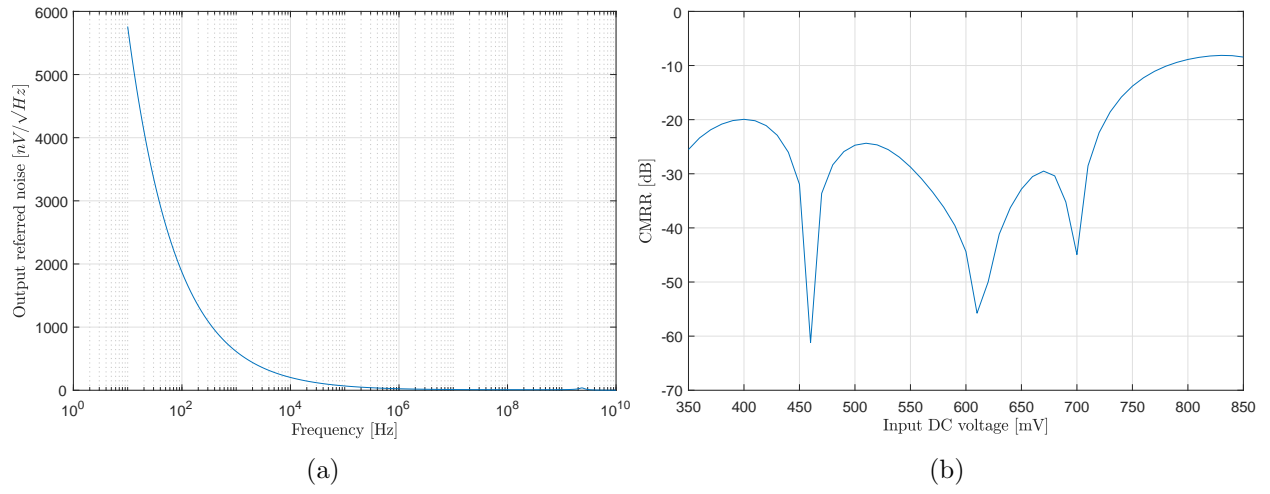


Figure 5.12: (a) Sampling cell buffer noise spectrum. (b) Sampling cell buffer common mode rejection ratio.

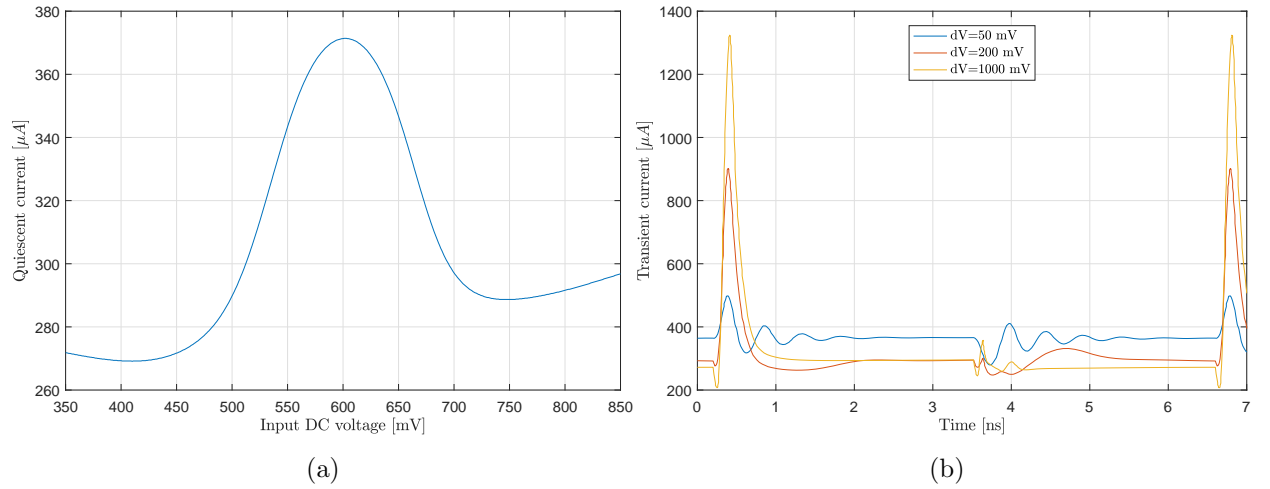


Figure 5.13: (a) Sampling cell buffer quiescent current draw as a function of input voltage. (b) Sampling cell buffer transient current draw for three different output voltage amplitude swings.

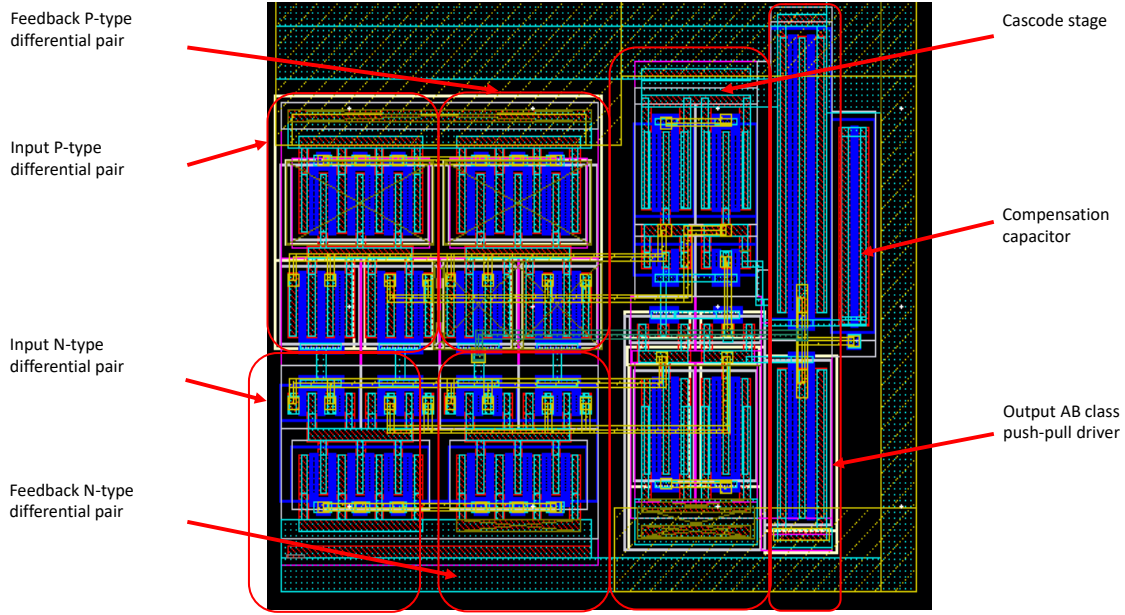


Figure 5.14: Sampling cell buffer layout.

5.2.3 Differential Clock Driver

The differential clock driver is composed of two fully differential logic AND gates whose core circuit is realized with a cross-coupled quad. The advantage of the cross-coupled quad is its ability to deskew differential signals up to a couple hundred ps. The drawback is the inherent propagation delay. Figure 5.15 shows the propagation delay of the differential clock driver as a function of temperature. The added jitter is on the order of a few femtoseconds, thus presenting a negligible contribution to the total added jitter of the array.

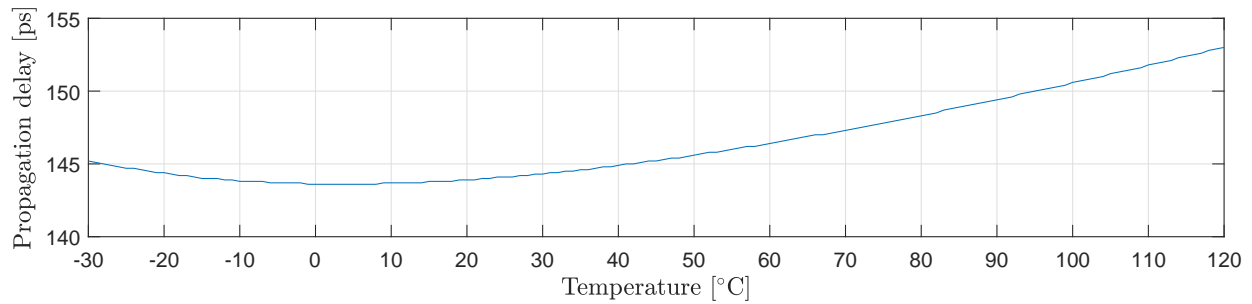


Figure 5.15: Differential clock driver propagation delay as a function of temperature.

5.2.4 Summary of the Switched Capacitor Array

The switched capacitor cell simplified schematic is shown in Figure 5.16a. Each sampling cell is driven by one differential AND gate. The timing diagram of the signals driving the two sampling cells is shown in Figure 5.16b. The strobe signals of blocks A (QA, nQA) and B (QB, nQB) are shifted in time by 3.2 ns in respect to each other, thus ensuring that the two sampling blocks are never concurrently sampling and transferring.

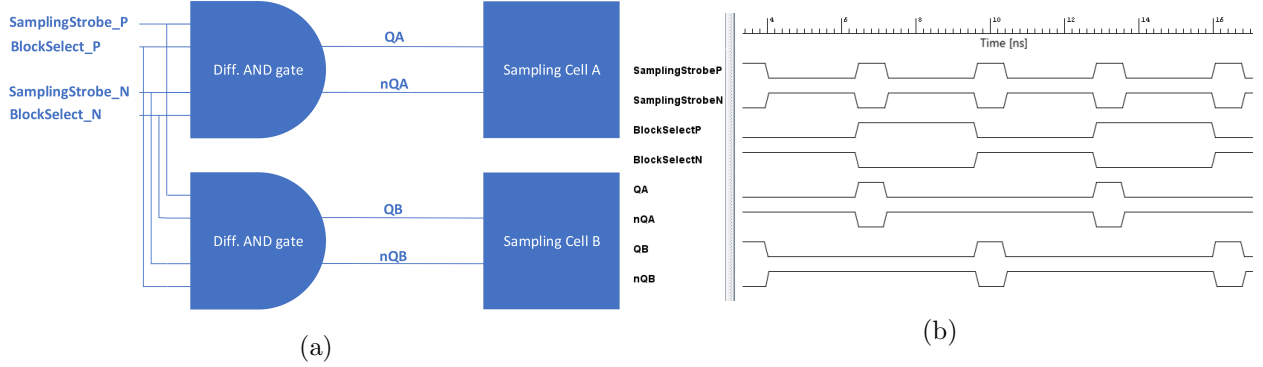


Figure 5.16: (a) Simplified schematic of the switched capacitor cell. (b) Timing diagram of the switched capacitor cell strobe signals.

The layout of the switched capacitor cell is shown in Figure 5.17. The size of the cell is $51.47\text{ }\mu\text{m} \times 21.13\text{ }\mu\text{m}$.

Figure 5.18a shows the transient response of the sampling cell. During the switch transition from track-mode to hold-mode, the charge under the switch transistors' gates is displaced towards both sides of the switch. The charge injected towards the transmission line results in a transient response, which is absorbed by the input channel buffer. Conversely, the charge injected towards the sampling capacitor adds to the present charge and thus results in a DC voltage offset that is ultimately amplitude dependent. By considering the original signal amplitude versus the reconstructed sampled signal amplitude, the amplitude dependent offset voltage turns into a virtual gain of the sampling cell. Combined with the differential to single-ended conversion, this results in a virtual gain of the switched capacitor cell of 2.13 V/V . Unfortunately, some gain non-linearity is still present. Figure 5.18b shows the reconstructed sampled signal amplitude, the input signal amplitude, a linear fit, and the associated gain non-linearity as functions of the single-ended signal amplitude. The dominant factor contributing to the overall gain non-linearity is the non-linear response of the sampling cell buffer.

The overall dimensions of the switched capacitor array are $1310.74\text{ }\mu\text{m} \times 51.47\text{ }\mu\text{m}$. Furthermore, the SCA present a capacitive load of approximately 2.1 pF to the input channel buffer.

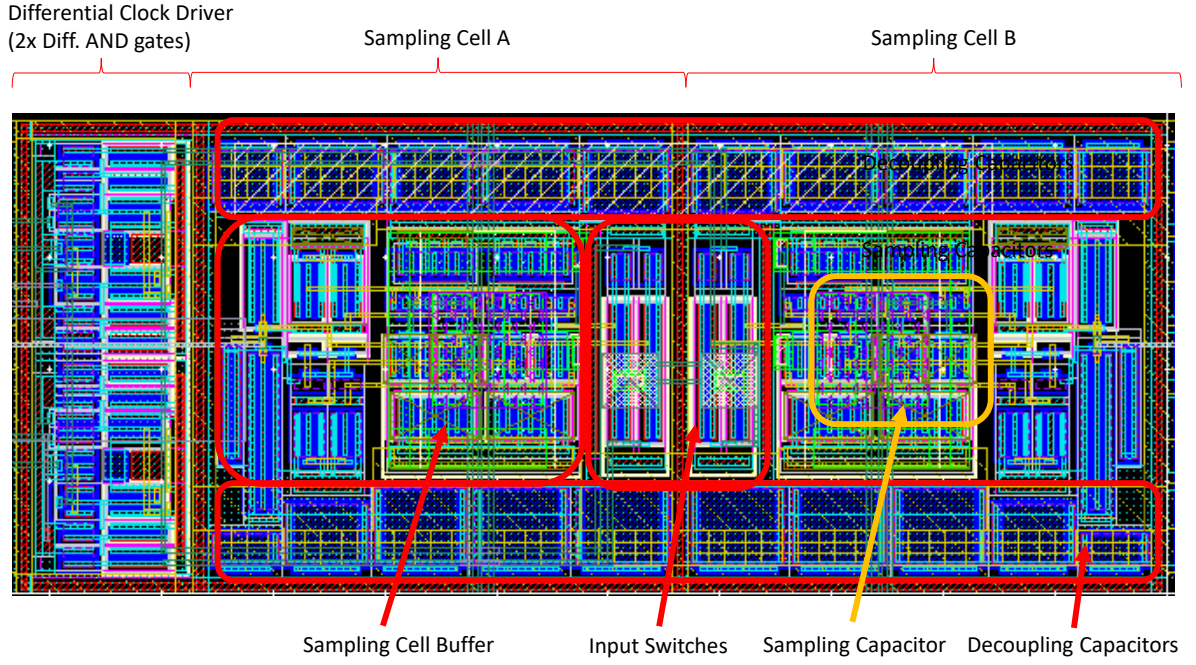


Figure 5.17: Switched capacitor cell layout.

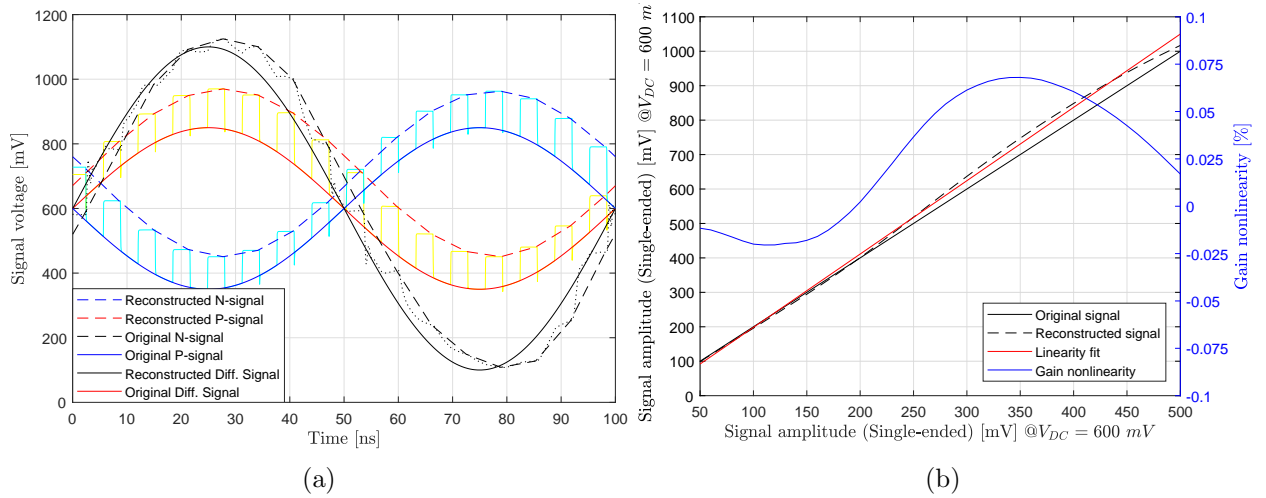


Figure 5.18: (a) Sampled single-ended sampling cell signals with the corresponding reconstructed signals. (b) Original and reconstructed signals as functions of single-ended input signal amplitude.

5.3 Analog Storage Array

Analog buffering is used for storing the voltage samples awaiting digitization. Such a scheme has been chosen for the following reasons:

- Information is stored on a capacitor, thus considerably increasing information density as opposed to digital storage;
- The analog storage cell does not need refreshing, hence, no power is drawn by a cell in hold mode.

The drawback of this scheme is the constant degradation of the stored voltage with time.

Each channel has its own analog storage array, which is composed of 2048 storage cells. The array is arranged in a row-column format containing 32 rows and 64 columns. The 32 rows are further split into two blocks of 16 rows, each corresponding to one of the SCA blocks. To minimize the capacitive load, each sampling cell is connected to 16 storage cells in parallel. That is, one sampling cell is connected to one column of the storage array block as it is shown in Figure 5.19.

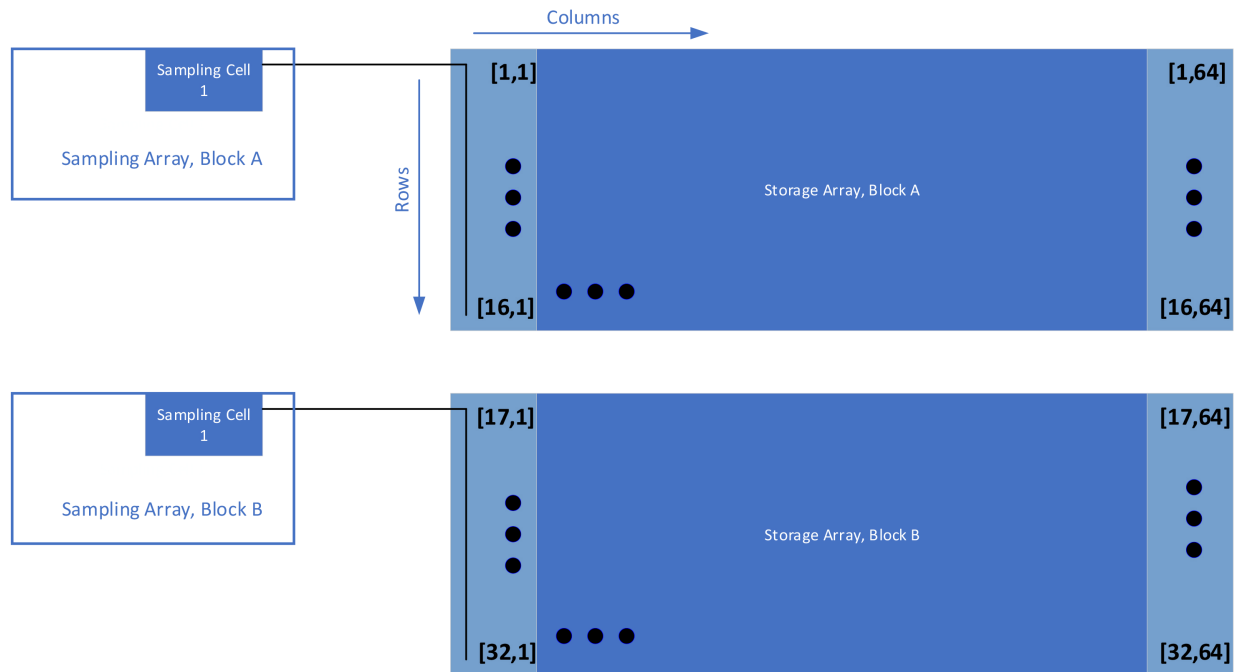


Figure 5.19: Storage array functional schematic portraying the row-column format of the storage cells.

Each storage cell is composed of the following parts:

- input switch;

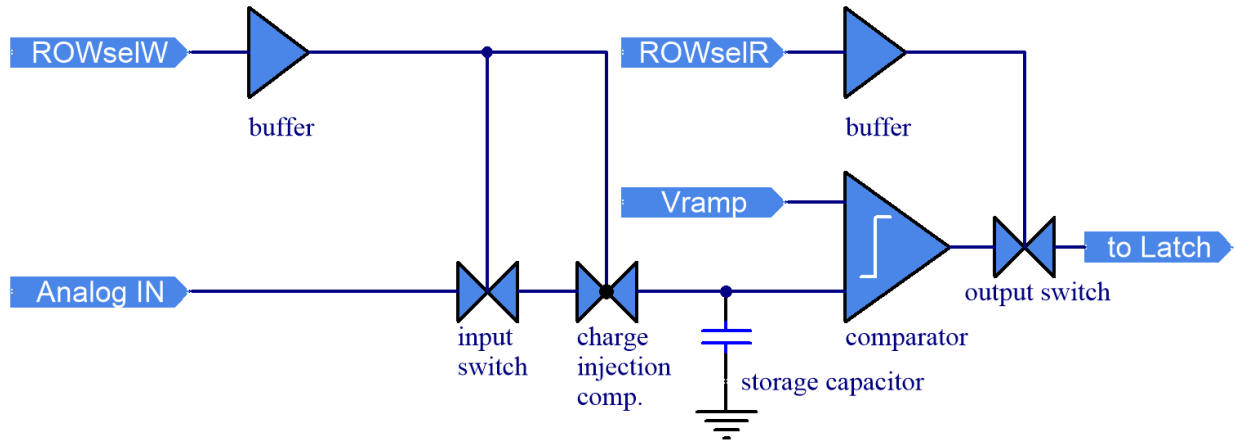


Figure 5.20: Storage cell functional schematic.

- charge injection compensation;
- storage capacitor;
- comparator;
- output switch.

5.3.1 Analog Storage Cell

Figure 5.20 shows the functional schematic of the analog storage cell.

The input switch is composed of two complementary CMOS transistors. To maintain a stable sample voltage for up to $5\ \mu s$, the leakage current of the switch needs to be minimized. This is accomplished by using high-threshold transistors. The current implementation has a worst case voltage degradation of $143\ \mu V/\mu s$ or $714\ \mu V$ per $5\ \mu s$.

The charge injection compensation is achieved by connecting a complementary CMOS switch in series with the input switch, while at the same time driving it with the opposite phase. That is, when the input switch is opening, the compensation switch is closing and vice versa. This way the charge that gets injected towards the storage capacitor gets collected by the compensation switch. The compensation switch has the drains and sources of the transistors connected together so as to not attenuate the signal coming through.

The storage capacitor is implemented with transistors as opposed to capacitors provided by the process. The advantage is a higher capacitance per unit area. The disadvantage is the degraded linearity. That is, the capacitance changes over the dynamic range. To keep the capacitance variation as small as possible, two complementary CMOS high-threshold transistors with long lengths have been used. Both transistors have their gates connected to the input switch, while their

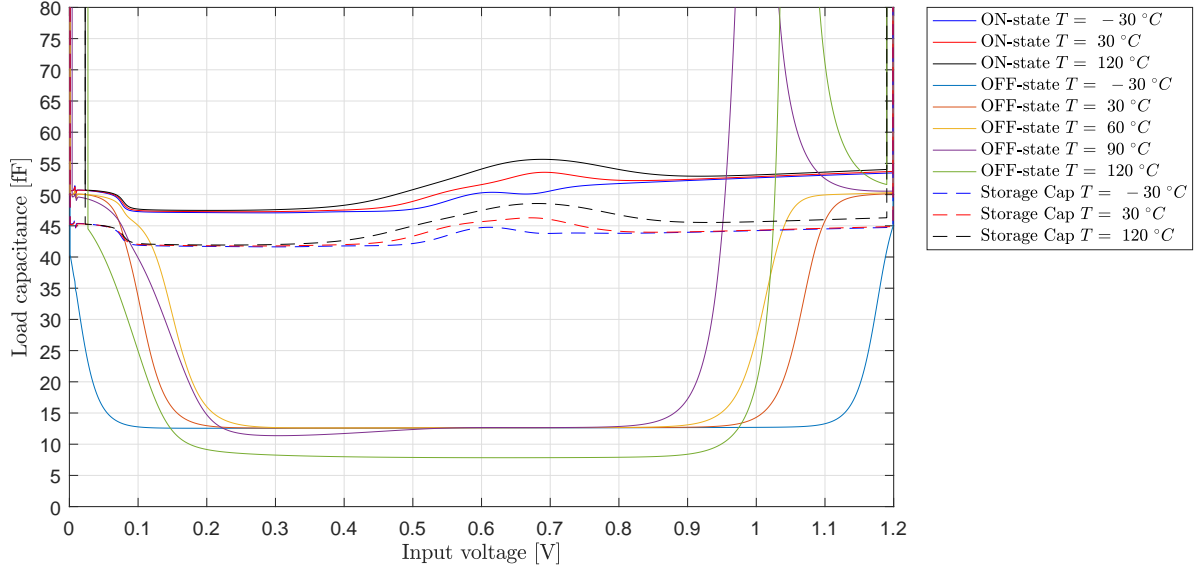


Figure 5.21: Storage cell ON-state, OFF-state, and storage capacitances as functions of input voltage at various temperatures.

sources and drains are connected to the supply rails. The NMOS drain and source are connected to the ground rail, while the PMOS drain and source are connected to the positive rail. Figure 5.21 shows the simulation results for the load and the storage capacitances as functions of the input voltage at various temperatures. The storage and the ON-state capacitances are sufficiently linear throughout the entire voltage and temperature ranges. The OFF-state load capacitance of the switch shows a currently unexplained behavior at high input voltages and temperatures. However, as long as the operating temperature remains below $60\text{ }^{\circ}\text{C}$, this behavior is acceptable. In summary, the average input load capacitance in the ON-state ranges from 47 fF to 53 fF , while the input load capacitance in the OFF-state is approximately 12.7 fF . The extrapolated storage capacitance ranges from 41 fF to 46 fF .

Figures 5.22a and 5.22b, respectively show the switch ON-state resistance and the storage cell tracking bandwidth as functions of input voltage at different temperatures. Both, the resistance and the tracking bandwidth, exhibit considerable variation in terms of the input voltage. However, the cell has been designed to maintain a minimum bandwidth of 345 MHz at $30\text{ }^{\circ}\text{C}$, thus satisfying the transfer speed of 312.5 MHz .

With such a high variance of the bandwidth, distortion is expected. Figures 5.23a and 5.23b respectively show total harmonic distortion as a function of signal amplitude and signal frequency at different temperatures. The THD as a function of amplitude is the highest at the center of the dynamic range, where the bandwidth is the lowest. In addition, THD decreases with the increasing temperature and increases with increasing frequency.

The ON-state input referred noise of the storage cell is approximately $3.2\text{ nV}/\sqrt{\text{Hz}}$. For the

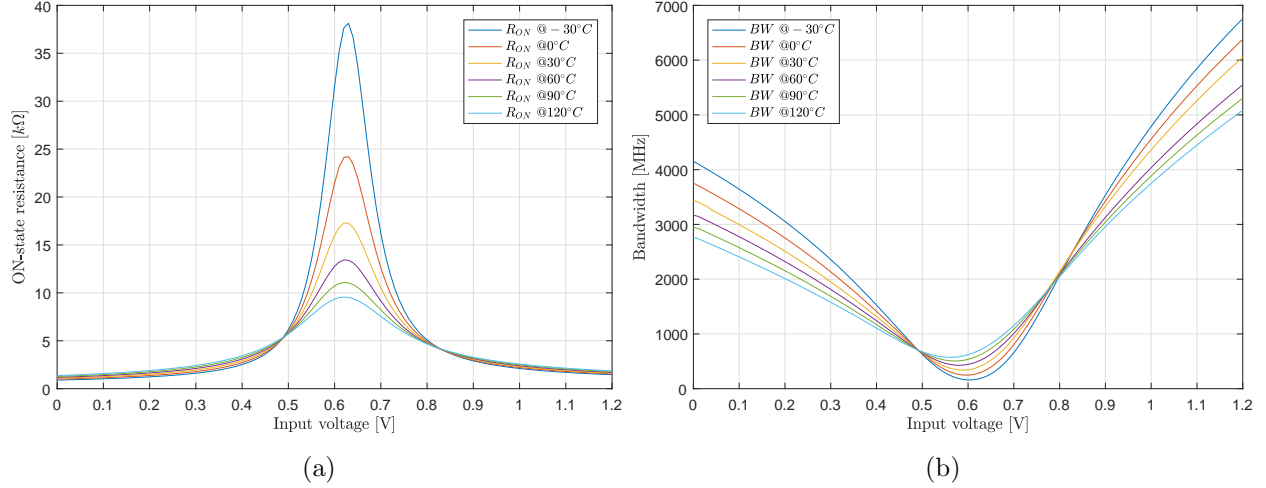


Figure 5.22: (a) Switch ON-state resistance as a function of input voltage at different temperatures. (b) Storage cell tracking as a function of input voltage at different temperatures.

given bandwidth, the integrated noise floor is approximately $75.6 \mu V$. As in the case of the sampling cell, the OFF-state noise floor is dominated by the kTC noise. Considering the storage capacitance of $41 fF$, the kTC noise term amounts to $318 \mu V$.

Calculating both, SINAD and ENOB, we found that the noise floor allows for an ENOB of $11.83 bits$.

The storage cell comparator is integrated in the storage cell, even though it is a part of the single-slope ADC. It is primarily designed to have a low form factor, low current, and a flat propagation delay response over the dynamic range. Its architecture is based on a N-type differential pair with an output class A source follower, which is in turn followed by a regenerative inverter. The static current draw of the comparator is approximately $1.3 \mu A$, while the propagation delay through the dynamic range is fairly flat with an average value of $12.3 ns$ as shown in Figure 5.24a. The standard deviation of the propagation delay is on the order of $2.86 ns$ at the temperature of $30^\circ C$. During digitization, the storage cell comparator serves as trigger for latching a high-speed counter value, thus providing the digital value in counts of the sampled voltage. The comparator is driven by a linear voltage ramp.

To keep the effects of quantization error low enough the digitizer requires a minimum of $9.7 bits$ [22]. However, in order to allow for some margin for error, it has been decided that a $12 bit$ vertical resolution is a good compromise between complexity and performance.

The propagation delay results in a voltage variable offset that directly translates into a sampling error. In the RFPix1 case, one ADC count is $489 ps$ or $244 \mu V$. Additional sources of error that need to be taken into account are the threshold variation of the comparator and the charge injection from the storage cell input switch. Figure 5.24b shows all three sources of error and their sum as functions of the storage cell voltage. Despite the charge injection compensation circuit, the error

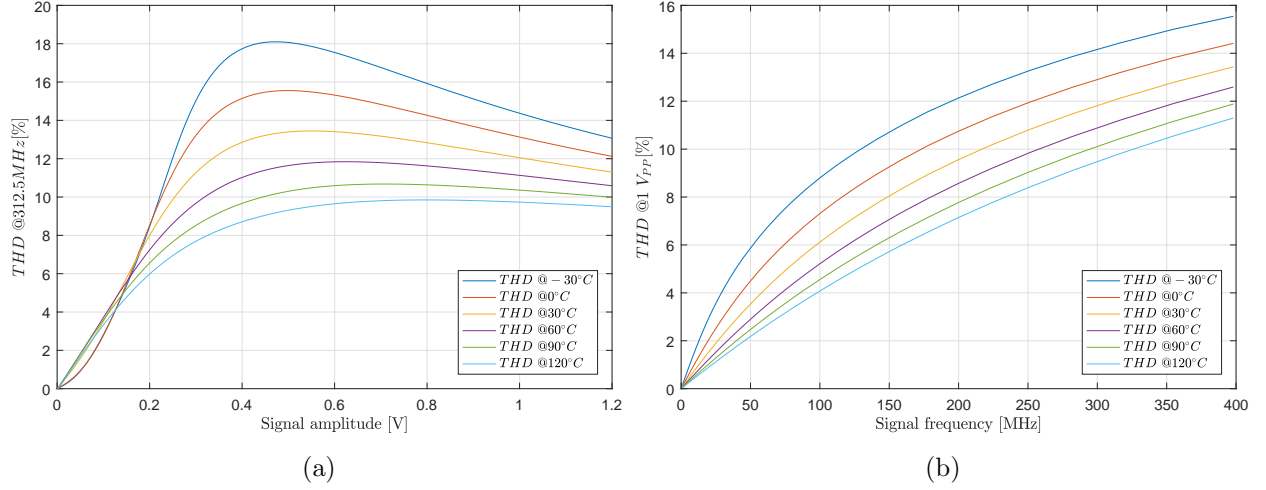


Figure 5.23: (a) Total harmonic distortion as a function of signal amplitude at different temperatures and signal frequency of 312.5 MHz. (b) Total harmonic distortion as a function of frequency at the signal amplitude of 1 V_{PP} .

originating with the charge injection still dominates. Unfortunately, charge injection is not a linear process. The compensation switch has been scaled to minimize the effect of charge around the center of the dynamic range, where the time of arrival is measured.

All the storage cells in a column are connected together to a common rail that connects to a single latch. Only one storage cell can be digitized at the time. To switch between cells, a pass gate has been added in series to the comparator output.

The storage cell layout is shown in Figure 5.25. The size of the cell is $20.47 \mu m \times 4.56 \mu m$.

A single storage array block present a load of approximately 222 fF to the sampling cell buffer.

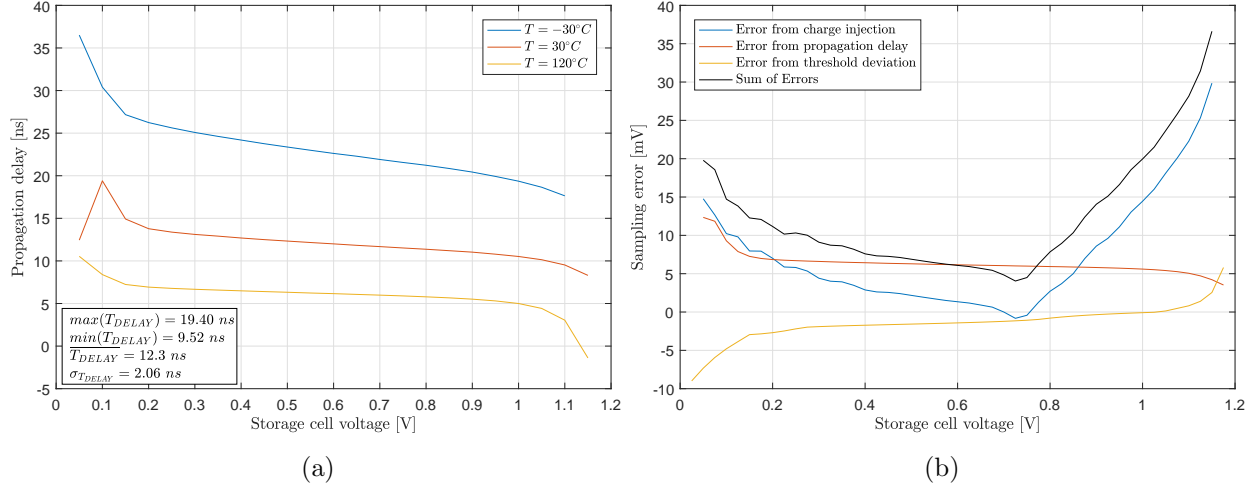


Figure 5.24: (a) Comparator propagation delay as a function of storage cell voltage at three different temperatures. (b) Cumulative storage cell error (from charge injection, propagation delay, and threshold variation) as a function of the storage cell voltage.

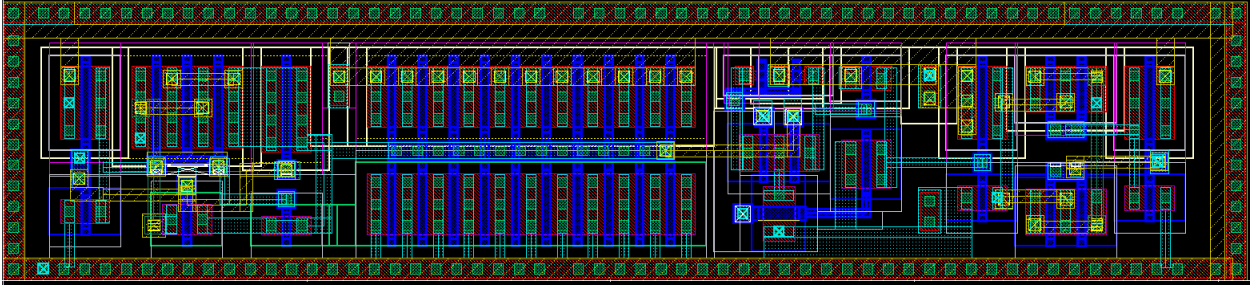


Figure 5.25: Storage cell layout.

5.3.2 RC Wire Model

The wires connecting the sampling and storage arrays provide some additional load capacitance to the sampling cell buffer. Figure 5.26a shows the capacitance and resistance as functions of line length for four line widths. The wires are laid out on layers five and seven with shielding metal on layers three and six. By simulating various wire geometries, it has been determined that the wire resistance decreases faster than the wire capacitance increases, thus supporting the case of wider wires. However, computing the bandwidth as shown in Figure 5.26b indicates that the bandwidth for the longest and narrowest wire is still above 5.5 GHz . To save space and minimize the capacitive load towards the sampling cell buffer, the wires are kept at their minimum widths. Consequently, the wire capacitance is approximately 88 fF . The total load capacitance seen by the sampling cell buffer output is then 310 fF .

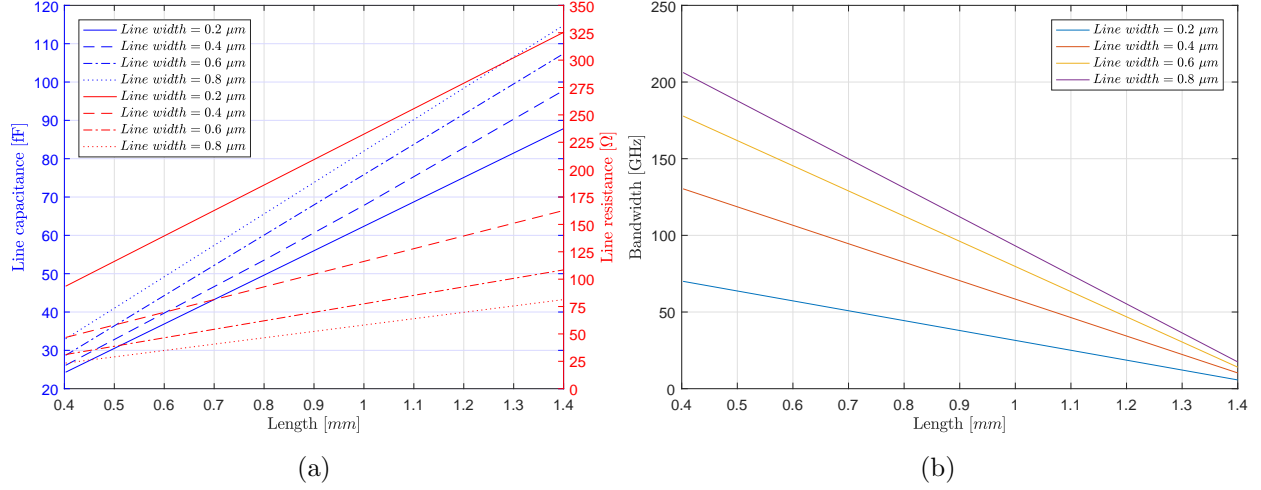


Figure 5.26: (a) Capacitance and resistance for the wire connecting the sampling and storage arrays as functions of length for four different wire widths (b) Bandwidth of the wire connecting the sampling and storage arrays as a function of length for four different wire widths.

5.3.3 Analog Buffer Depth Estimation

The depth of the storage array has been determined by simulating a RFPix1 digitization model, which included the Belle II event model, the estimated detector efficiency, and the ASIC readout model. The entire algorithm for estimating the buffer depth has been coded in MATLAB[®].

The simulation assumes a time window of 2 ns , which is the minimum longitudinal spacing in time of the electron bunches in the SuperKEKB accelerator ring. Hence, it is also the minimum time interval between two collisions, which represent the events. Every event results in a number of particles that can hit the detector. In addition, the detector is experiencing background hits as discussed in chapter two. Consequently, there are two mechanisms that need to be modeled and run in parallel:

- The trigger rate of the Belle II acquisition system, which generates digitization triggers when an event of interest is detected;
- The detector hit rate which includes the actual hits results from the event, the background hits, and the detector dark count hits.

The average Belle II trigger rate ($\mu_{TriggerRate}$) is 30 kHz , while the average detector hit rate ($\mu_{HitRate}$) has been estimated in chapter two and is approximately $7.1 \cdot 10^8 \text{ Hits}/\text{cm}^2 \cdot s$. Both represent the average constant rate of randomly occurring events and thus can be modeled using the Poisson distribution [73]. The probability of an event happening a certain number of times within a time window is given by Equation 5.2,

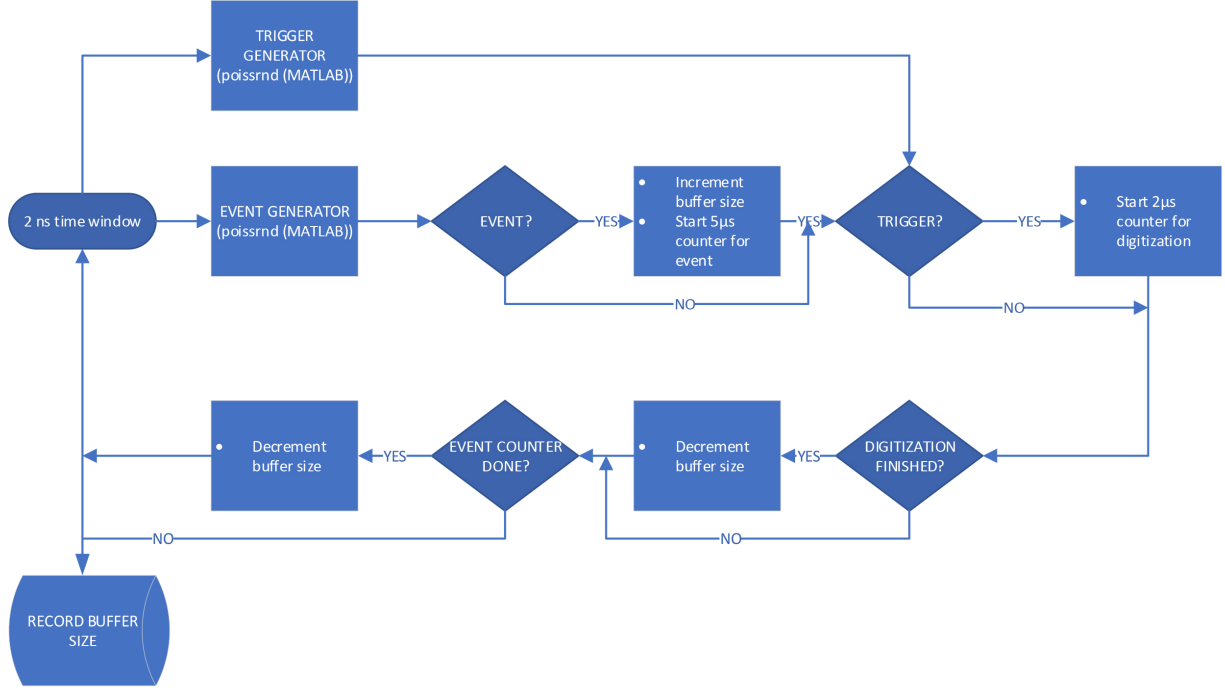


Figure 5.27: Buffer depth simulation flow chart.

$$P(n \text{ events in time window}) = e^{-\lambda} \frac{\lambda^n}{n!} \quad (5.2)$$

where n is the number of events occurring in one time window (2 ns), e is the Euler's number (2.71828...), and λ is the average number of events per time window. The λ for the trigger rate is calculated using Equation 5.3,

$$\lambda_{TRIGGER} = \overline{TriggerRate} \cdot dt = 6 \cdot 10^{-5} \quad (5.3)$$

where dt is the time window used in the simulation. The λ for the detector hit rate is calculated using Equation 5.4,

$$\lambda_{HITS} = \frac{\overline{HitRate}}{E[\max(\rho_{hitrate})]} = 5.7 \cdot 10^{-3} \quad (5.4)$$

where $E[\max(\rho_{hitrate})]$ is the saturated hit rate density expectation value of the detector defined in chapter two.

In order to account for trigger latency in the system, the buffer needs to retain the sampled voltage for $5 \mu\text{s}$. In addition, the ASIC needs some time to perform the digitization of the sampled voltage. Currently the ASIC is being designed to accommodate digitization times ranging from $0.5 \mu\text{s}$ to $2 \mu\text{s}$. Figure 5.27 shows the simulation flow chart.

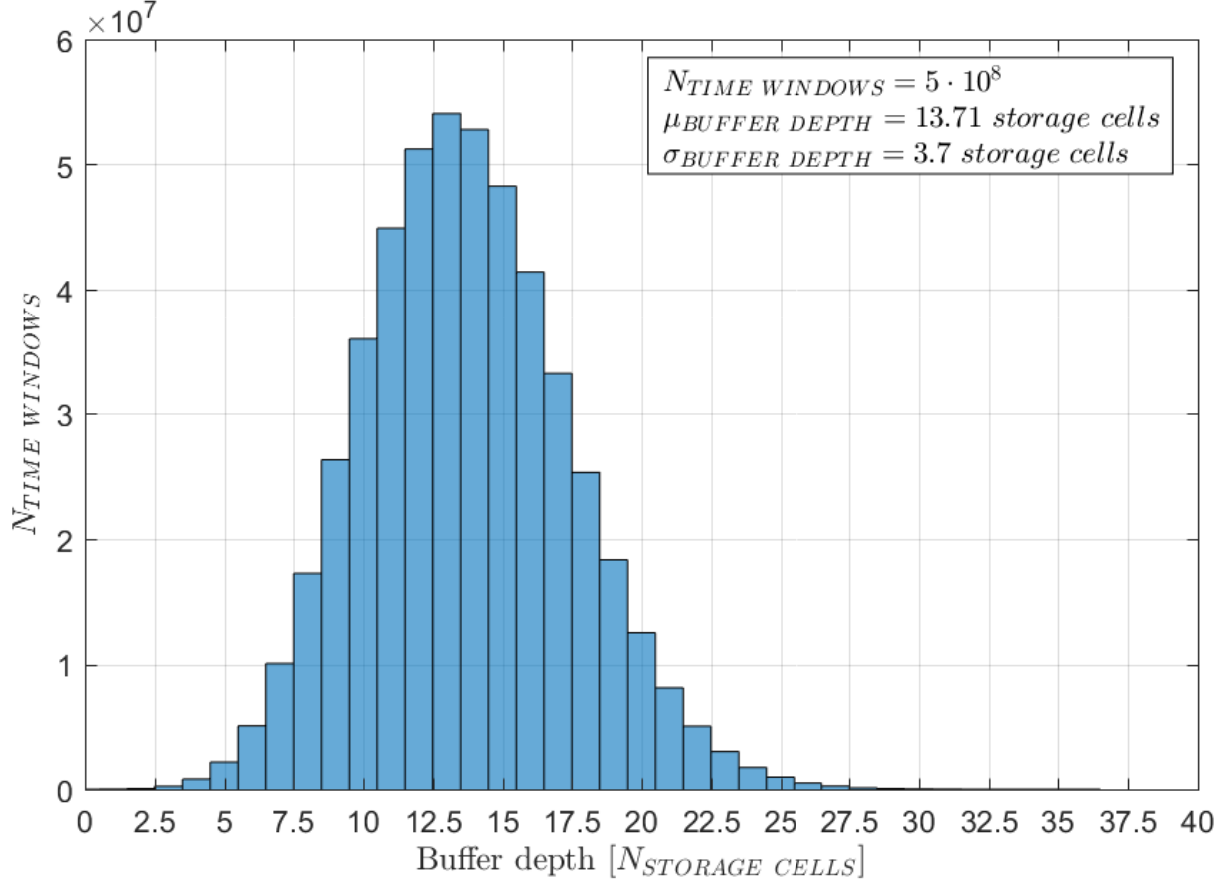


Figure 5.28: RFpix1 analog storage histogrammed buffer depth per time window.

The simulation has been run for one full second or $500 \cdot 10^8$ time windows. Figure 5.28 shows the histogrammed buffer depth per time window. The mean buffer depth is 14 cells, while the maximum buffer depth required to cover all of the events is 37 cells. However, the number of time windows needing a buffer depth higher than 32 is only 0.0013 %. It has been decided that all of these time windows maybe treated as a small efficiency loss, thus implementing a buffer depth of 32.

5.3.4 Summary of Analog Storage Array

The size of the entire storage array is $134.97\ \mu m \times 1287.085\ \mu m$, thus fitting into the channel spacing of $180\ \mu m$. The current draw of the entire array is approximately $2.46\ mA$.

5.4 Two-Level Delay-Locked Loop

The DLL provides the switched capacitor array with the sampling cell strobe (SCS) signals, which drive the sampling cell switches. Figure 5.29 shows a simplified high level schematic of a DLL. The basic principle of operation of every DLL is:

1. The input clock is routed to the delay line as well as the phase detector;
2. The clock signal propagates through the delay line with every delay element adding a deterministic delay to the signal;
3. At the output of the line, the clock signal gets fed back to the phase detector;
4. The phase detector is sensitive to differences in phase between the source clock and the delayed clock;
5. The phase detector generates a DC voltage that is proportional to the phase difference between the two clocks;
6. This voltage controls the delay elements, thus fine-tuning them in order to phase lock the two clock signals. This two clocks are delayed by one full clock cycle in reference to each other.

The outputs of the delay elements are used as strobe signals (SCS) for the sampling array. In a phase locked state, each delay element will add a delay given by Equation 5.5.

$$t_{DELAY} = \frac{T_{CLOCK_PERIOD}}{N_{DLL_CELLS}} \quad (5.5)$$

5.4.1 Jitter and its Propagation in Inverters and D Flip-Flops

Inverter Jitter

In our case, the most crucial parameter of the DLL is the added jitter. The advantage of a DLL over a phase-locked loop (PLL) is that the added jitter is not recirculated through the feedback, thus avoiding accumulation of the jitter [74]. The obvious disadvantage is the lack of flexibility in terms of frequency generation [49], where the DLL operates only at its reference frequency, while a PLL can generate multiples of the reference clock, depending on the ratio of the phase detector dividers.

The most common delay element in DLLs is the inverter. Jitter in inverters is a consequence of the amplitude noise being converted into phase noise (AM-to-PM modulation). It has been shown that the thermal noise of the channel resistance is increasing with scaling and is expected to further increase if mobility enhancement techniques are used. Also, noise amplification may start to significantly contribute to noise accumulation in a delay line for devices under 130 nm as

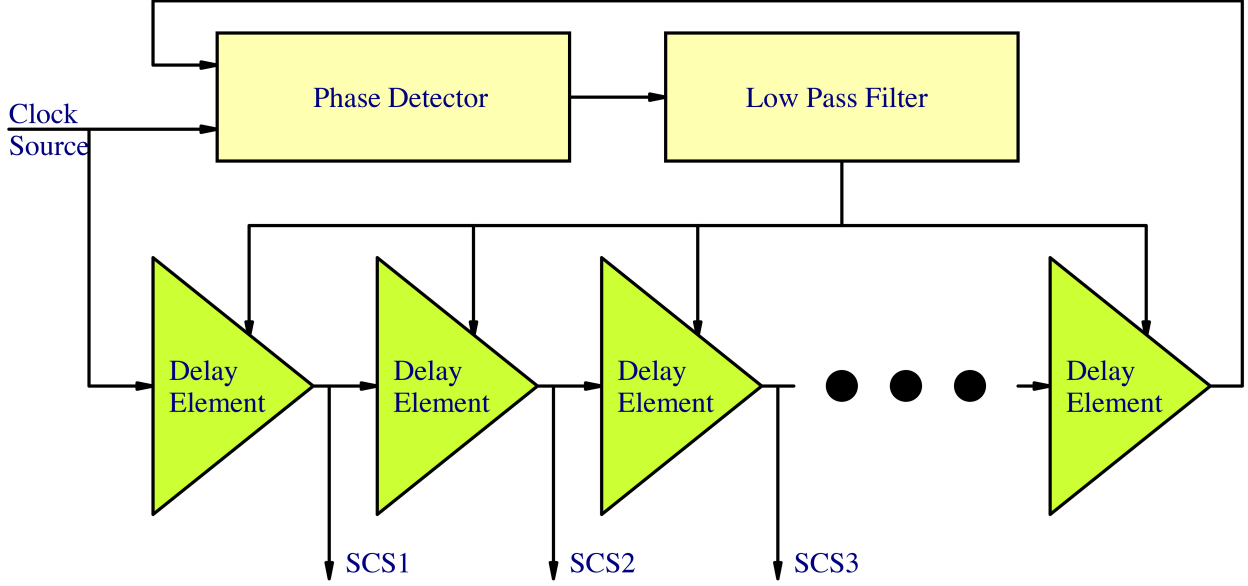


Figure 5.29: Functional block schematic of a DLL.

it depends essentially on the transconductance gain. On the other hand, supply noise coupling to the gates output has been shown to be decreasing with scaling (although this may be different for technologies under 130 nm , with higher junction capacitances), as well as the saturation current sensitivity to its value [71].

Figures 5.30a and 5.30b show the results of a jitter simulation for a single inverter at frequencies of 156 MHz and 312.5 MHz as functions of input clock rise time and current consumption. The jitter was calculated by integrating the simulated phase noise over the frequency range of 10 Hz to 10 MHz . Furthermore, Figure 5.31 shows added jitter as a function of temperature.

Simulation results indicate that clock frequency ($156\text{ MHz} - 312.5\text{ MHz}$) and rise time ($2\text{ ps} - 272\text{ ps}$) have minor impacts on the added jitter of the inverter. Conversely, the incremental current draw seems to be directly correlated with the scaling of the added jitter. Furthermore, we notice that current draw depends on the transistor sizes or, more accurately, on their drain to source resistance (R_{DS}). Thus, we conclude that the majority of the inverter added jitter originates from the thermal noise of R_{DS} as discussed in [71]. In addition, temperature dependence of added jitter in case of the inverter has been found to be approximately $0.019\text{ fs}/^\circ\text{C}$.

It is expected that jitter accumulates with the increasing number of delay elements along a delay line. Figures 5.32a and 5.32b show the added jitter and rise time as functions of the number of delay elements connected in series for four rise time values of the input clock. Rise time at the output of the delay elements settles to $\approx 31\text{ ps}$ after the first four delay elements even at the worst case input rise time of 272 ps (every inverter is also a regenerative amplifier). Due to this fact, the accumulation of jitter does not follow precisely the geometric series ($\sqrt{(N)} \cdot n$), where N is the

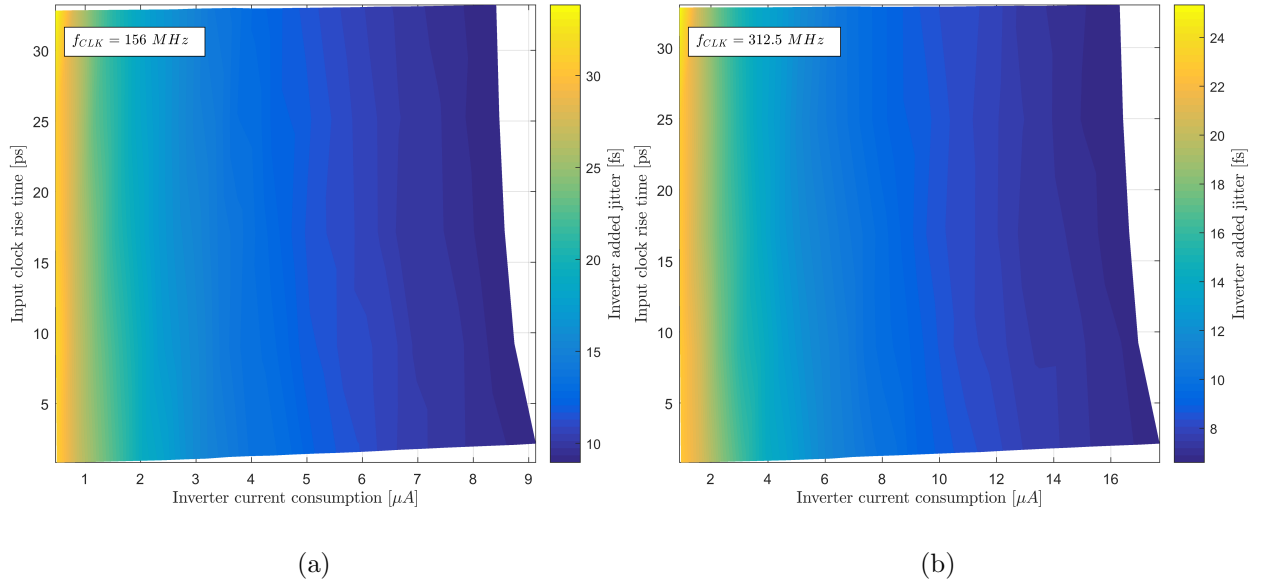


Figure 5.30: Added jitter of a single inverter as a function of input rise time and current consumption for a clock frequency of 156 MHz (a) and 312.5 MHz (b).

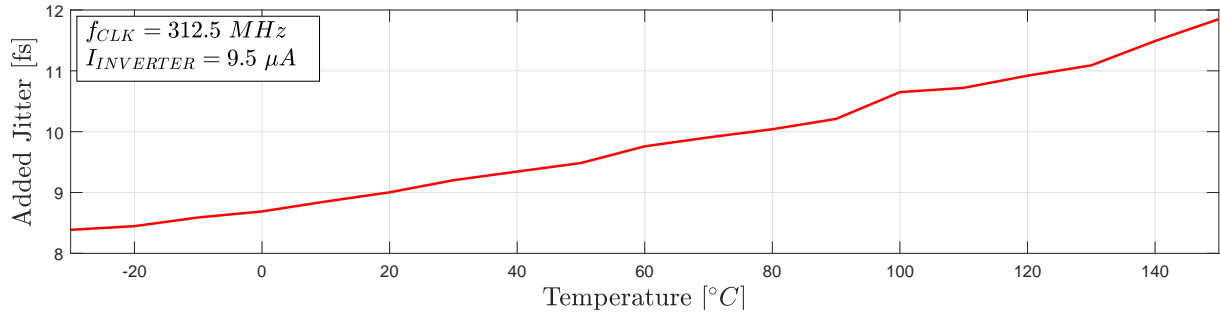


Figure 5.31: Inverter added jitter as a function of temperature.

number of delay elements in series and n is the jitter of a single inverter.

D Flip-Flop Jitter

D flip-flops can be realized in many different ways. We designed and studied a fully differential D flip-flop (DDFFC) that is based on transmission gates. Figure 5.33a shows the simplified schematic of the DDFFC.

The two basic building blocks are the inverter and the transmission gate. Therefore we expected the added jitter of the DDFFC to be the sum of the inverter contributions along the signal path from the input D to the output Q of the DDFFC. However, that seems not to be the case. To understand this, we need to observe what happens when the output of the DDFFC changes logic state. The update to the output logic level happens on the rising edge of the clock. At that

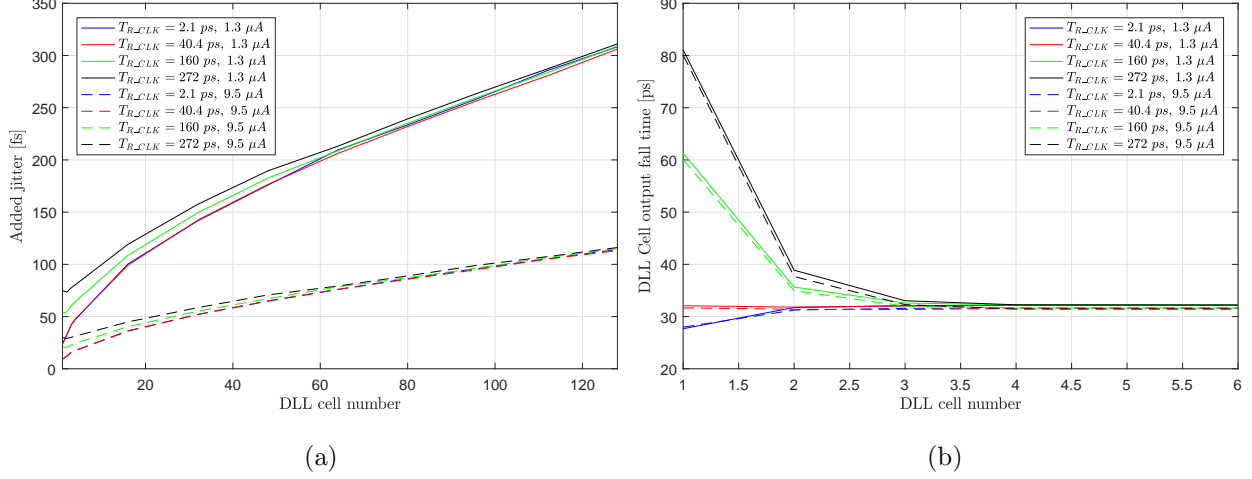


Figure 5.32: Added jitter (a) and rise time (b) along the delay line as functions of DLL cell number.

point, the input of the output inverters is already set, thus there is no timing ambiguity being propagated from the input of the DDFFC. That is, the jitter that is fed into the input D of the DDFFC does not propagate to the output Q. Conversely, the logic state at the output Q changes when the output transmission gates switch on and off, thus the jitter present on the source clock is propagated to the output. Therefore, the added jitter of the DDFFC is the sum of the jitter contributions of the output inverters, output transmission gates, and the source clock. Figure 5.33b shows the simulated added jitter as a function of temperature at the outputs (Q_X, \overline{Q}_X) of three DDFFCs that are connected in series. The added jitter of the DDFFCs is at the same level for all three DDFFCs even though they are cascaded, thus confirming that the jitter does not propagate from one DDFFC to the next.

5.4.2 Two-Level DLL

Using a multilevel topology of the DLL increases the complexity of the circuit and its power consumption. However, as seen in the previous section, added jitter grows with the increasing number of delay cells, even when operating at higher current. In a 128 sampling cell structure, the added jitter grows above the target 100 fs . Thus, reducing the number of sampling cells from 128 to 64 reduces the added jitter to about 77 fs . Furthermore, temperature also affects the added jitter. Figure 5.34a shows the added jitter as a function of temperature and the number of DLL cells. The added jitter scales almost linearly with the temperature from about 70 fs at $T = -30^\circ\text{C}$ to 89 fs at $T = 120^\circ\text{C}$. Figure 5.34b shows the jitter margin to 100 fs as a function of added jitter. It has to be noted that the DLL jitter and the clock source jitter have been taken into account. The jitter margin for 64 delay cells in series at $T = 120^\circ\text{C}$ and $f_{CLK} = 312.5 \text{ MHz}$ is equal to approximately 61 fs .

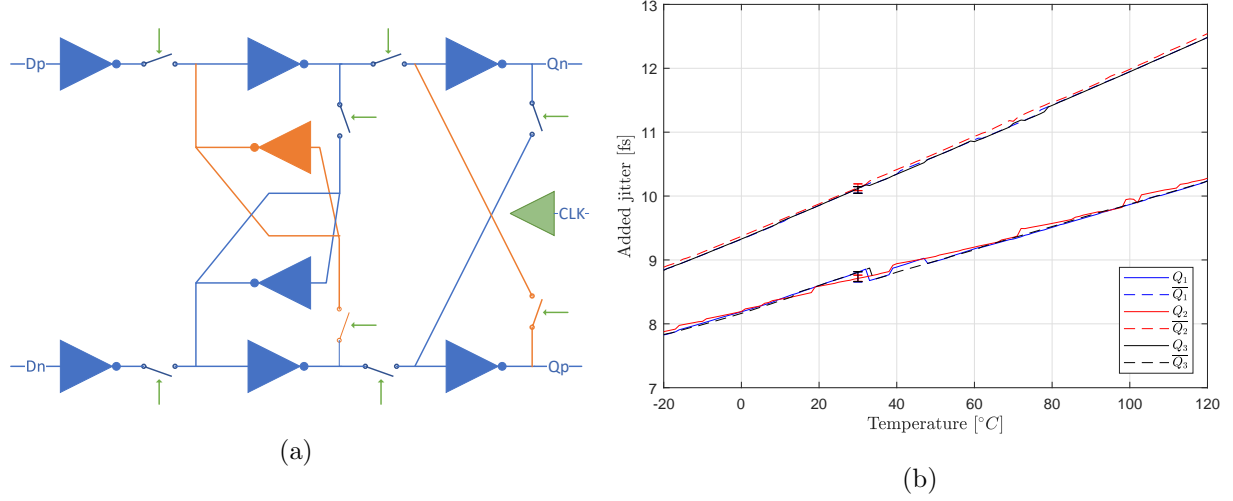


Figure 5.33: (a) Simplified block schematic of the modified fully differential D flip-flop. (b) Added jitter as a function of temperature at the outputs of three cascaded DDFFCs.

Considering that we did not take into account power supply noise, clock buffer jitter, and input channel buffer jitter, this margin needs to be increased as much as possible. In terms of added jitter, the best option would be to use a single array of DDFFCs. However, the obvious drawback is the fact that the source clock is required to run at the native sampling speed of the SCA (20 GHz), which is impractical and difficult to implement. Furthermore, the design would most likely result in a substantial increase in power consumption. One way of reducing the added jitter is to reduce its accumulation along the DLL. By using a multilevel topology where a short first level delay line (N delay elements) is used to drive several short second level delay lines (M delay elements), the accumulation of timing ambiguity is greatly reduced. Several multilevel architectures ($N \times M$) have been explored in terms of added jitter and power consumption. Figure 5.35 shows the comparison between the four architectures based on inverters only: one-level DLL (1×64), two-level DLL (4×16), two-level DLL (8×8), and a three-level DLL ($4 \times 4 \times 4$).

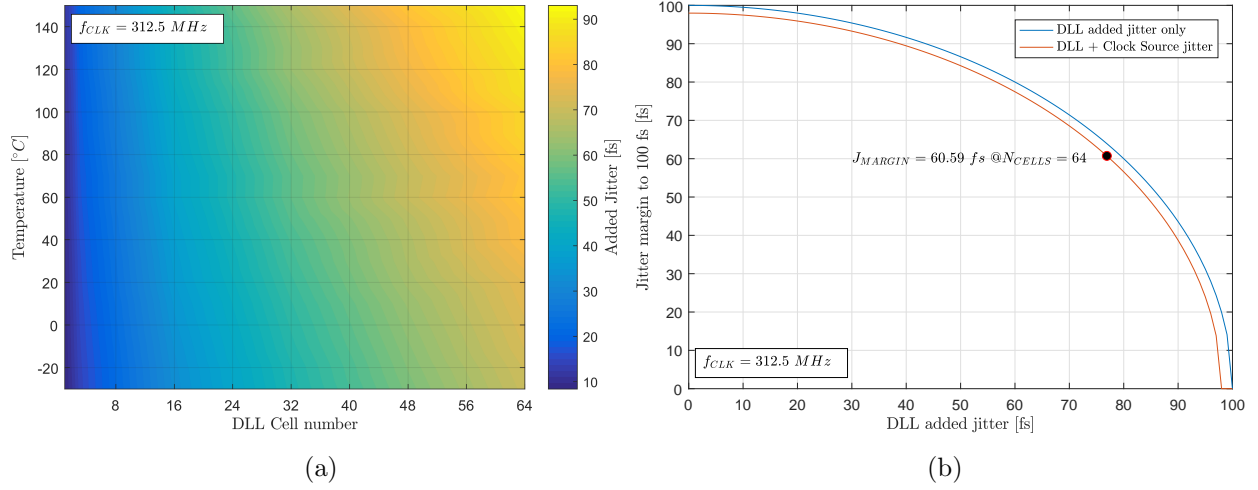


Figure 5.34: (a) Added jitter as a function of DLL cell number and temperature (b) Jitter margin to 100 fs as a function of added jitter.

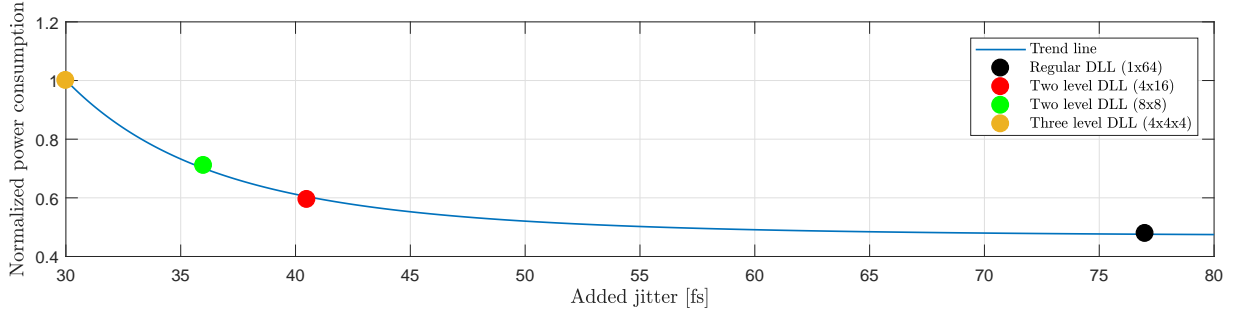


Figure 5.35: Comparison between DLL architectures in terms of power consumption and added jitter.

The best compromise between the added jitter and power consumption has been found to be a two-level DLL (4×16) topology. The simplified schematic of the designed two-level DLL circuit is shown in Figure 5.36.

There are three subcircuits to this architecture:

1. First level delay line (L1DL);
2. Pulse generation subcircuit (PGS);
3. Second level delay line (L2DL).

First Level Delay Line (L1DL)

The first level delay line is composed entirely of fully differential D flip-flops, while the second level delay line is composed of pseudo differential starved inverters. Each L2DL is driven by one L1DL

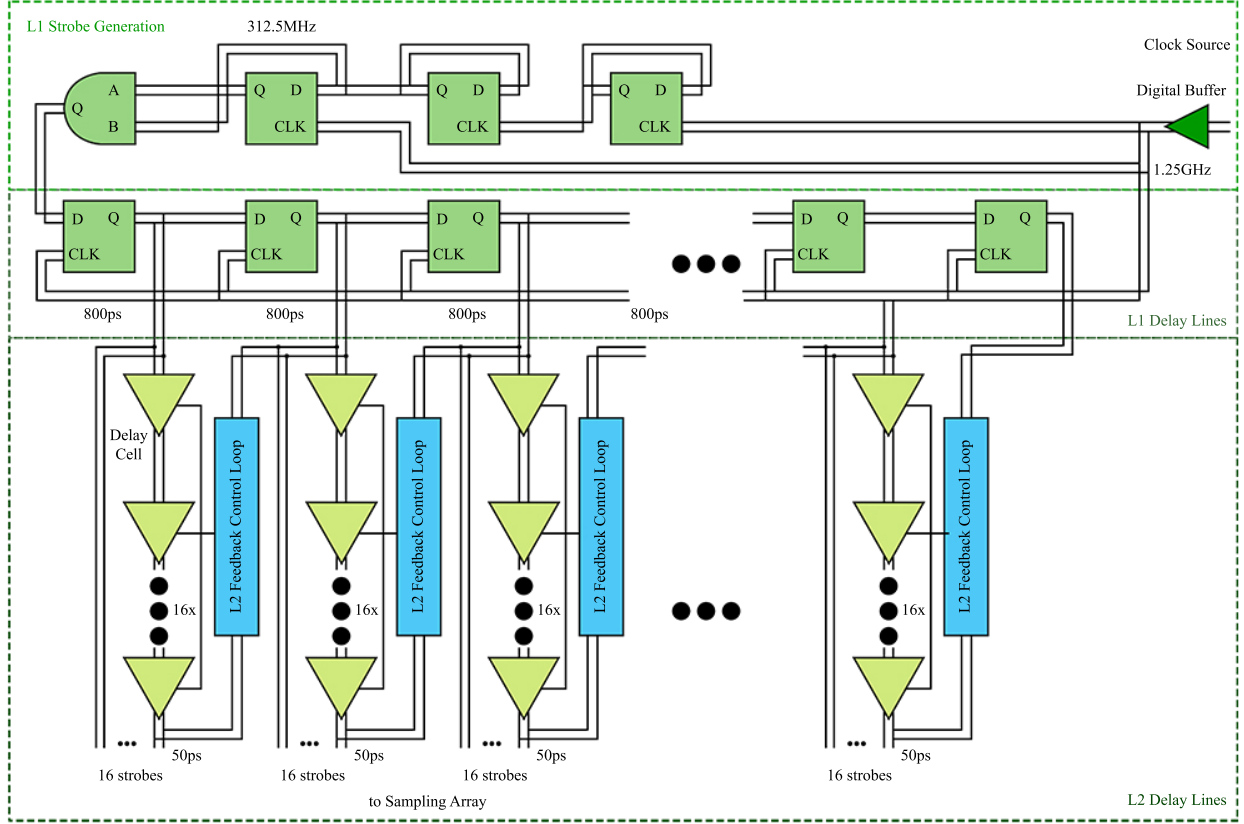


Figure 5.36: Simplified block schematic of a two-level cascaded DLL.

delay element. Consequently, the individual L2DL lines are not directly related, thus avoiding the accumulation of jitter from one L2DL line to the next. Each L2DL accumulates 800 ps of delay, thus, if properly spaced in time, four L2DL are sufficient to cover the required delay of 3.2 ns . To evade the overlap between the sampling windows, each of the L2DL lines needs to operate in the sampling mode for 800 ps while remaining in hold mode for 2.4 ns . Furthermore, only one of the L2DL lines can be in the sampling mode at the same time. To satisfy this requirement, a square wave, which is normally used in delay lines, has been replaced by a pulsed signal with a 25 % duty-cycle. The pulsed signal ensures that only one of the L2DL is in sampling mode at a given time. In addition, driving the DDFFCs of the L1DL line with the source clock (f_{CLK}) of 1.25 GHz provides with a propagation delay of 800 ps , thus ensuring the proper spacing in time of the L2DL lines. The pulsed signal is generated in the PSG subcircuit.

L1DL is composed of four cascaded DDFFCs. Figure 5.37a shows the cell to cell delay of the first two L1DL delay cells as a function of temperature including a $\pm\sigma$ mismatch estimate at 30 $^{\circ}C$. Due to a very low drift of the delay as a function of temperature and a very low impact of mismatch, it has been decided that a delay control loop is unnecessary. Crucial steps in achieving this were: the realization that the output of the DDFFCs is directly synchronized with the source clock and

the use of a symmetric layout of the DDFFC, where the source clock is buffered and routed from the output (Q) of the DDFFC towards the input (D), which ensured the minimum propagation delay from the source clock input (CLK) to the DDFFC output (Q). The layout of the DDFFC is shown in Figure 5.37b.

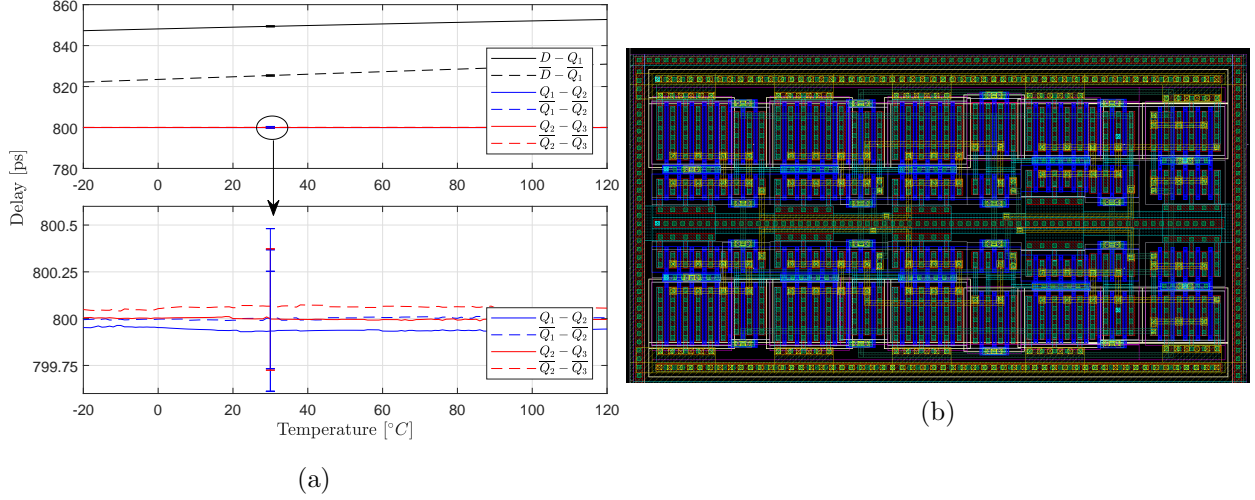


Figure 5.37: (a) L1DL cell to cell delay as a function of temperature with a $\pm\sigma$ mismatch estimate at 30 °C. (b) Layout of the modified RFpix1 DDFFC cell.

Pulse Generation Subcircuit (PGS)

The first part of the pulse generation subcircuit is designed to divide f_{CLK} by four, thus generating a clock signal (f_{DLL}) of 312.5 MHz, whose period of 3.2 ns exactly matches the overall delay of the DLL. The second part of the PGS decreases the f_{DLL} duty-cycle to 25 %, thus creating the required pulsed signal, which is then fed into the L1DL. It is worth noting that the PSG follows a fully differential configuration.

The source clock digital buffer is composed of two inverters and has an added jitter of 1.5 fs. The timing uncertainty due to mismatch is on the order of $\sigma = 62$ fs.

The block select strobe that is used to switch between the sampling block A and B is generated by further dividing the (f_{DLL}) by two to form a square wave with a frequency of 156.25 MHz. In addition, the block select strobe has delay element which enables the possibility of fine-tuning the alignment of the block select strobe with the sampling strobes. This same signal is routed to the analog storage array time base to generate the storage strobes.

The timing diagram of all the DLL signals is shown in Figure 5.38.

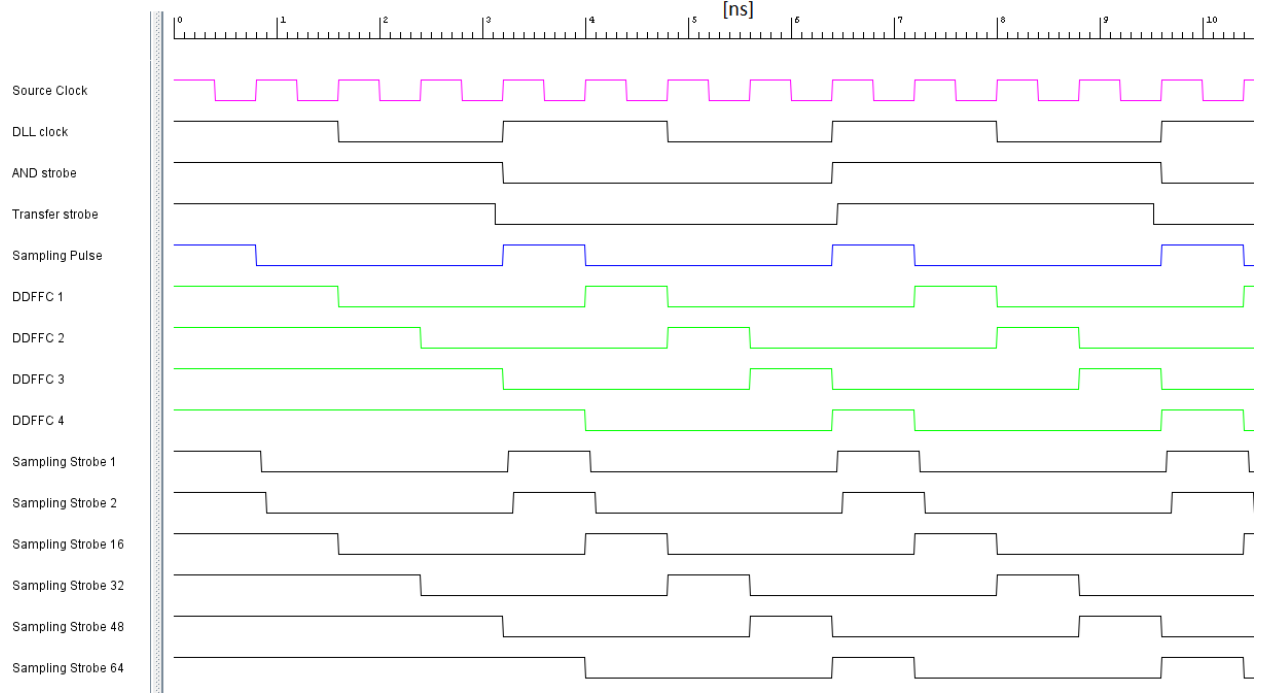


Figure 5.38: Timing diagram of the designed two-level DLL. The time scale used in the figure is in *ns*.

Second Level Delay Line (L2DL)

The L2DL is composed solely of pseudo differential starved inverters, due to the fact that the starved inverter is the only delay element that has been found to have a sufficiently low intrinsic input/output delay. In addition, each L2DL has its own dedicated feedback control that ensures the phase stability of the array.

Each pseudo differential starved inverter is composed of a pair of single-ended starved inverters. Special care has been taken in making the layout to minimize the effects of mismatch, temperature drift, and parasitics as shown in Figure 5.39.

Figure 5.40a shows the input/output delay of the starved inverter cell at the nominal delay of 50 *ps* as a function of temperature. The delay drifts from approximately 47 *ps* at -20°C to almost 55 *ps* at 120°C . Consequently, the temperature drift coefficients are:

- rising edge coefficient $K_{temp.Tr} = 55 \text{ fs}/^{\circ}\text{C}$;
- falling edge coefficient $K_{temp.Tf} = 52 \text{ fs}/^{\circ}\text{C}$.

Figure 5.40b shows the worst case added jitter of the L2DL line as functions of temperature for both the rising and the falling edges. It is worth noting that the jitter has been calculated differentially. To obtain the single-ended jitter, the differential jitter needs to be multiplied by $\sqrt{2}$.

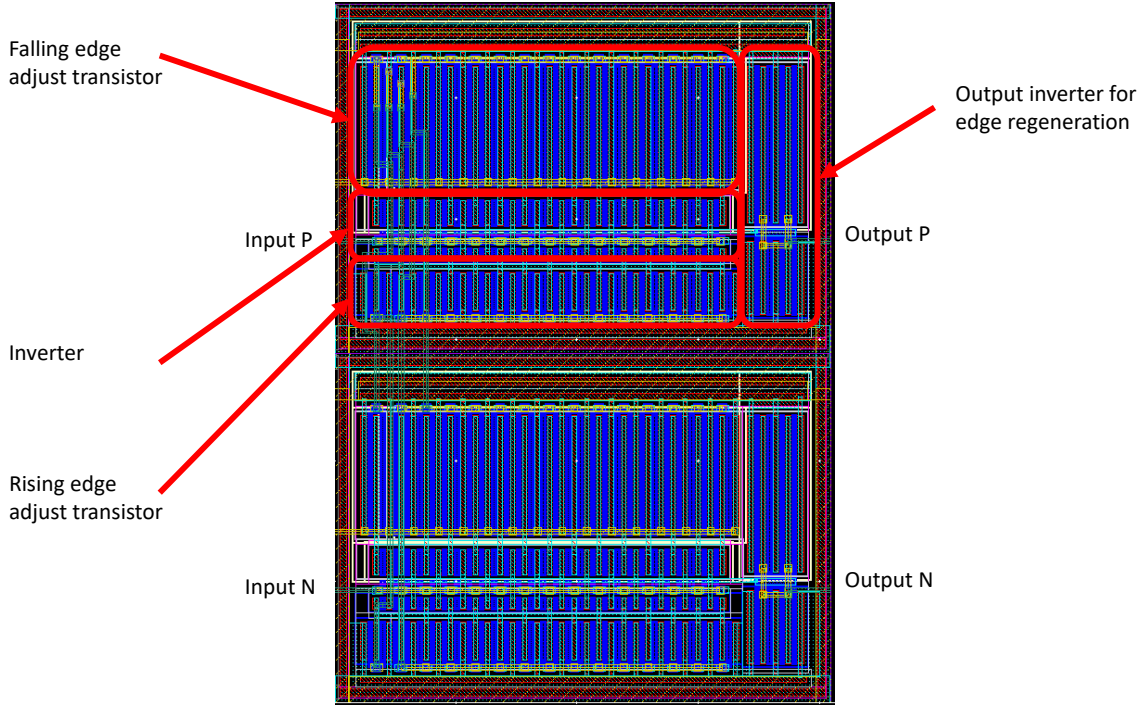


Figure 5.39: Pseudo differential starved inverter layout.

Figure 5.40c shows the cumulative delay deviation from 50 ps at 30 $^{\circ}C$ along the L2DL line caused by mismatch.

The starved inverters are designed to have an almost 90° conduction angle, thus making the adjusting of the input/output delay on the rising and falling edges almost independent of each other as shown in the Figure 5.41. Both edges are driven independently by the feedback control circuit.

L2DL Feedback Control

The L2DL feedback control subcircuit is composed of two independent control loops for both, the rising and falling edges, respectively. Figure 5.42 shows the block schematic of one such control loop.

Each of the control loops has two phase detectors, two charge pumps, one PI controller, and a linearization circuit that drives the L2DL line. Such an architecture is necessary to compensate for the phase detector error due to mismatch and temperature drift. Moreover, the phase detector and the charge pump are designed to have a non-zero current in steady state in order to avoid a dead zone in regulation and consequently a potential low frequency phase drift.

Both inputs of the reference phase detector are driven by the same clock signal, thus producing an integrated steady state voltage representing the locked condition. This signal is then used as the setpoint for the PI controller. The PI controller is realized with a single rail-to-rail operation

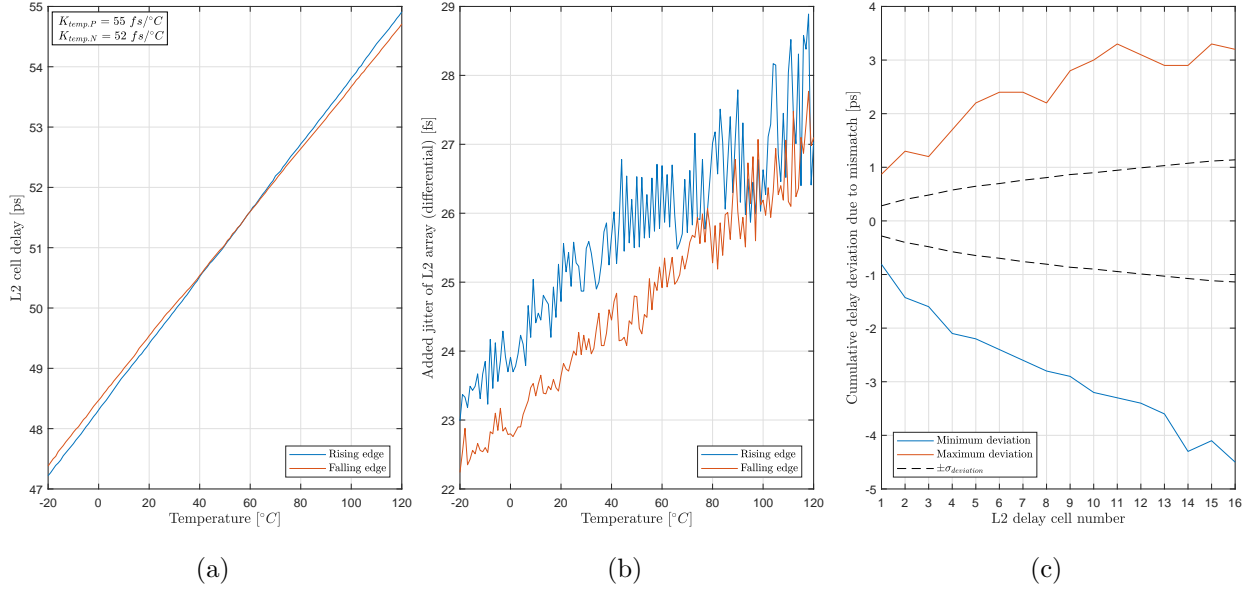


Figure 5.40: Simulated L2 cell delay (a), L2 delay line added jitter (b) as functions of temperature, and the cumulative delay deviation of the L2 delay line due to mismatch (c).

amplifier. The error signal of the controller is supplied by the feedback phase detector, which monitors the phase difference between the strobe signal from the last delay cell in the L2DL and the output of the DDFFC of the next L2DL line. The only remaining error of this architecture comes from the mismatch between the two phase detectors. To minimize this effect, both phase detectors are laid out in close proximity. Finally, the PI controller regulates the L2DL line control voltage through the linearization circuit in order to drive the error signal to match the setpoint, thus achieving a phased locked condition. The PI coefficients are: $K_P = 1$ and $T_I = 6 \cdot 10^{-6}$. The linearization circuit is designed to smoothen the L2DL response as shown in Figure 5.43, which proved necessary to avoid integral windup.

Figure 5.44 shows the phase error between the reference clock signal and the output of the L2DL on both, the rising and the falling edges, in time. The delay loops settle to a steady state phase error of approximately $\pm 0.2 \text{ ps}$ after approximately $17 \mu\text{s}$ or 340 sampling periods. Such a long integration time and consequently low loop bandwidth is important to keep possible phase disturbances outside of the bandwidth of interest. The low frequency cutoff of the bandwidth of interest is approximately 416 MHz [75].

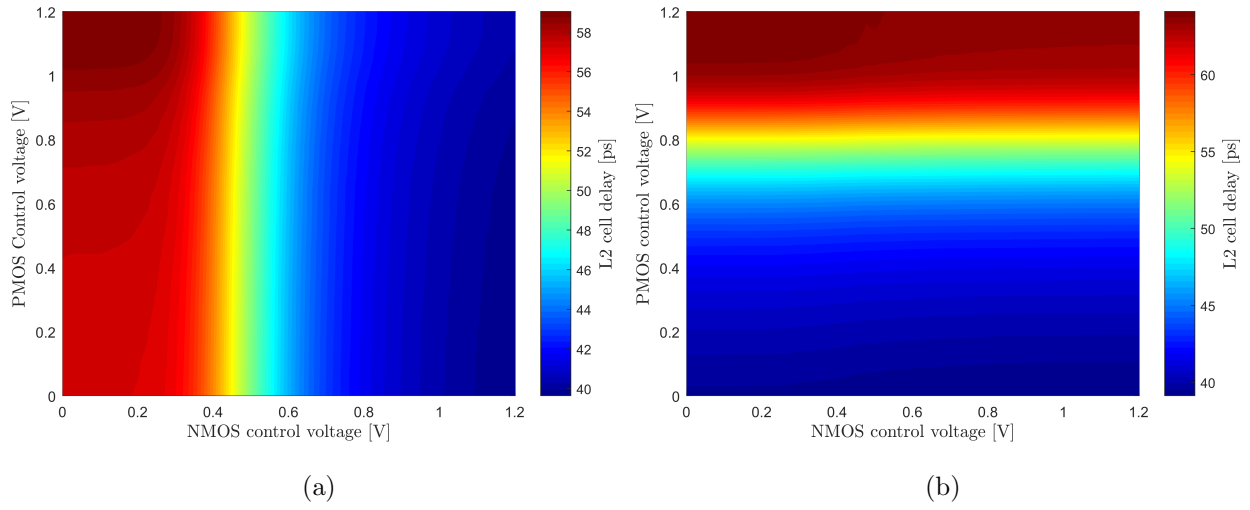


Figure 5.41: Simulated delay as a function of control voltages for both, the rising (a) and the falling edges (b).

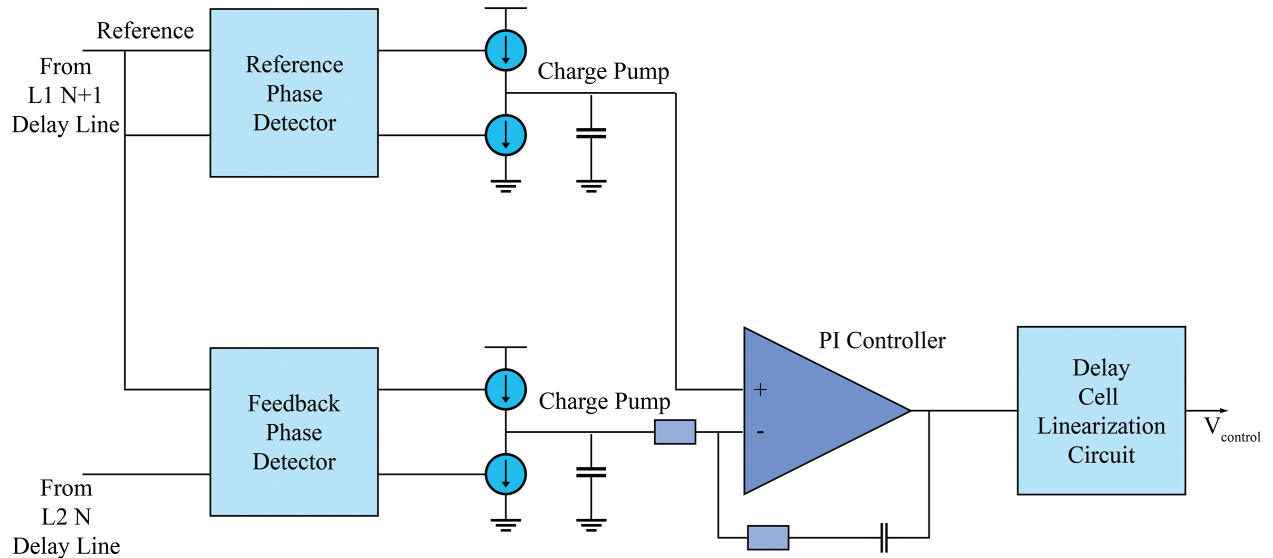


Figure 5.42: L2DL feedback control loop block schematic.

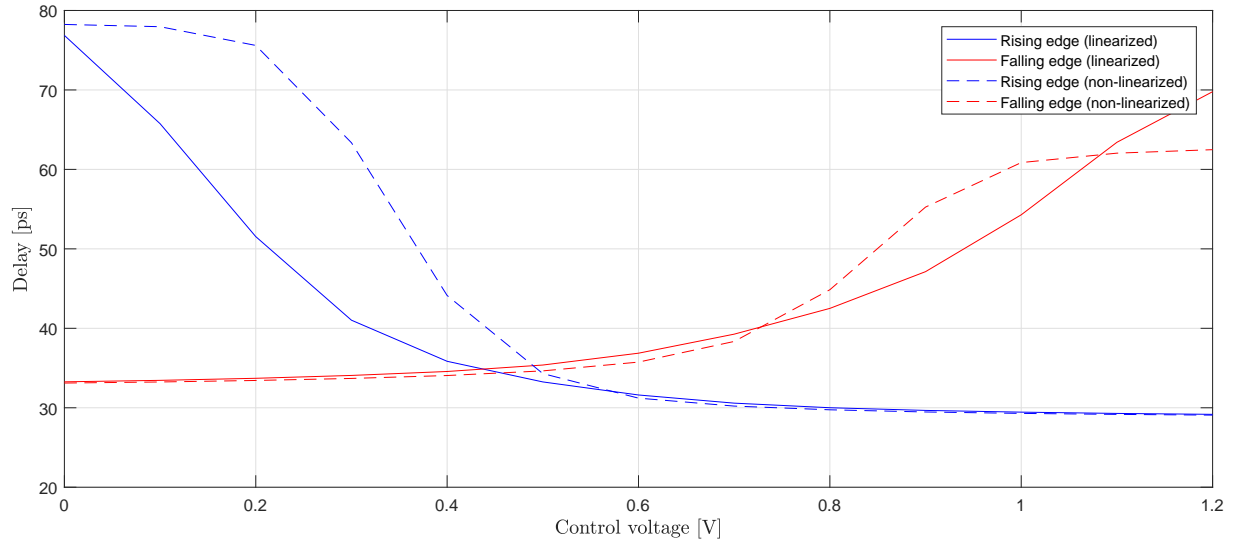


Figure 5.43: Smoothed L2DL response as a function of control voltage.

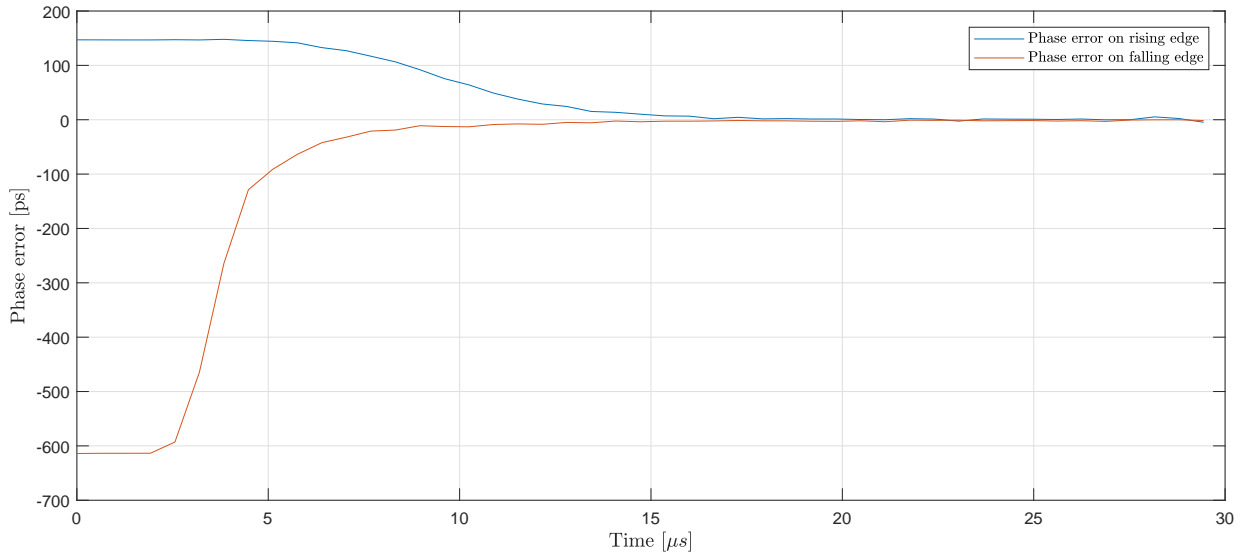


Figure 5.44: Phase error response of the L2DL array in time.

Two-Level DLL Summary

To decrease crosstalk and loading of the SCS lines, two channels share a single DLL. Each DLL is sourced by the input clock circuit.

The worst case total added jitter of the two-level DLL is simulated to be approximately 28 fs . The current architecture is designed to have 64 taps; however, it is possible to extend it to any number of second level delay lines with no degradation in terms of added jitter. Furthermore, the feedback control is always sourced from the local L2DL array and the adjacent DDFFC of the next L2DL, which are routed with equal distances in order to minimize the effect of a physical delay due to long lines. Figure 5.45 shows the two-level DLL layout with important parts highlighted.

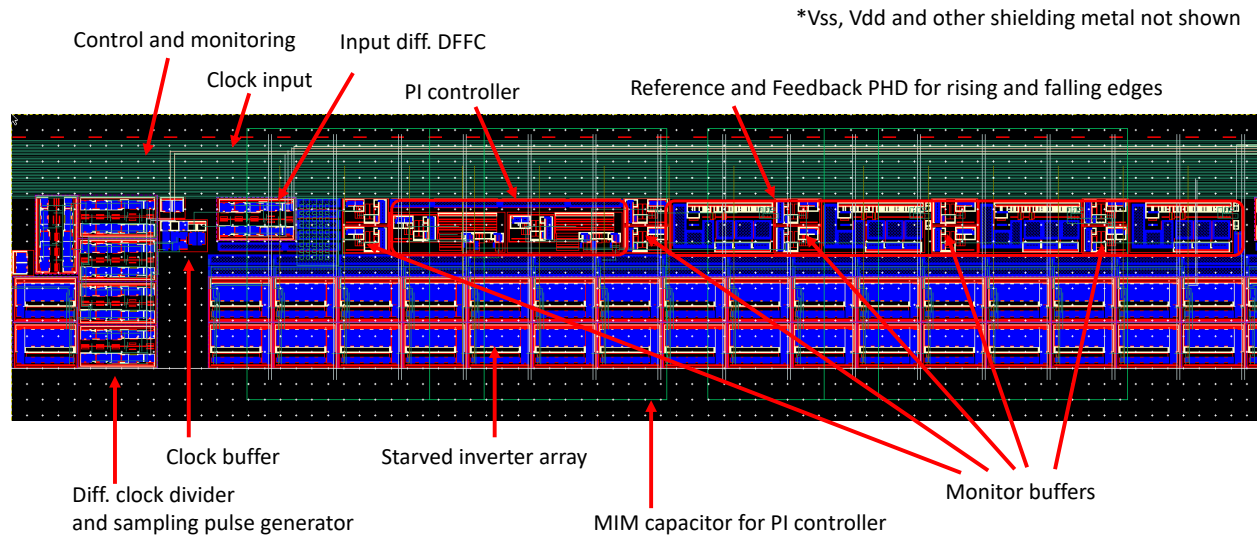


Figure 5.45: Snapshot of the two-level DLL with important parts highlighted.

The total current draw of the two-Level DLL is around 8.5 mA

5.5 General Purpose Input/Output Rail-to-Rail Operational Amplifier

Many analog signals inside the ASIC require precise buffering in order to decouple them from the loads and to isolate them from the disturbances that can be injected or coupled back from various parts of the chip. These signals can be:

- monitoring channels; the DLL circuit has 24 monitoring channels;
- DAC voltages (setpoints, bias voltages, etc.); The DLL has nine DAC signals.

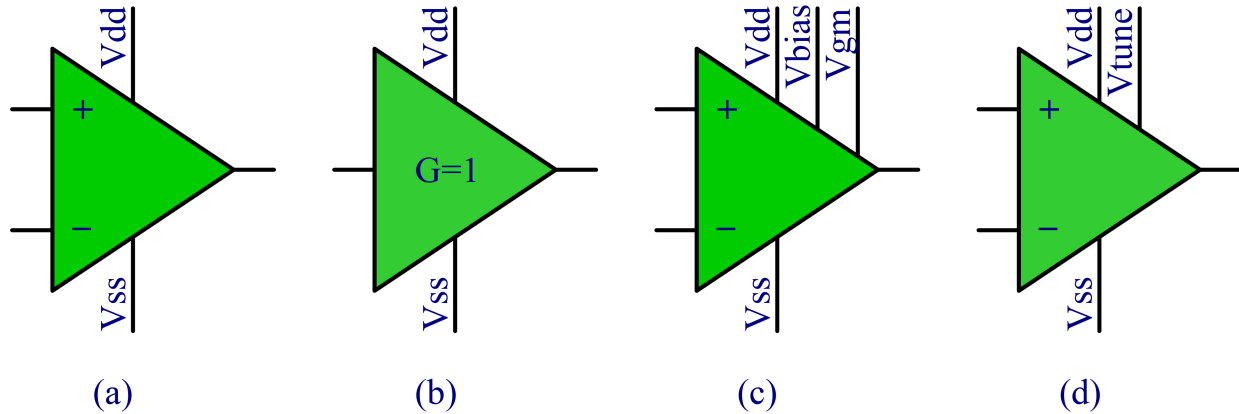


Figure 5.46: Schematic symbols of four variations of the general purpose input/output rail-to-rail operational amplifier implementation. (a) self-biased, (b) unity gain self-biased, (c) tunable biases (current source and g_m compensation), (d) tunable frequency response.

For this purpose, a set of general purpose op-amps has been designed. These amplifiers have the same core circuit, but feature different configurations depending on the application. Figure 5.46 shows symbols representing the four variations of the general purpose operational amplifier implementation.

All four have input and output dynamic ranges that span from the negative supply rail to the positive rail. The four variations include:

1. a self biased option: this avoids the need of DACs or any other biasing circuitry, which makes the design more compact;
2. a high speed option: The compensation circuit can be tuned with a DAC, which makes it possible to tune the frequency response of the amplifier;
3. an option for applications requiring higher precision: the bias voltages of the main current source and the g_m compensation circuit can be tuned externally to tune the DC response of the amplifier;
4. a self biased unity-gain option: the amplifier is already connected as a unity gain buffer and is thus ready to be directly implemented into a circuit.

Figures 5.47 and 5.48 show the transistor level schematic and the corresponding layout of the general purpose op-amp. The biasing stage has two circuits. The first one generates the bias voltage for the g_m linearization circuit, while the second generates the biases for both differential pairs. The main stage is composed of two complementary differential pairs with an integrated g_m linearization circuit. The core stage is followed by a folded cascode stage, which combines the currents from both differential pairs and provides additional open-loop gain. The cascode stage drives an output

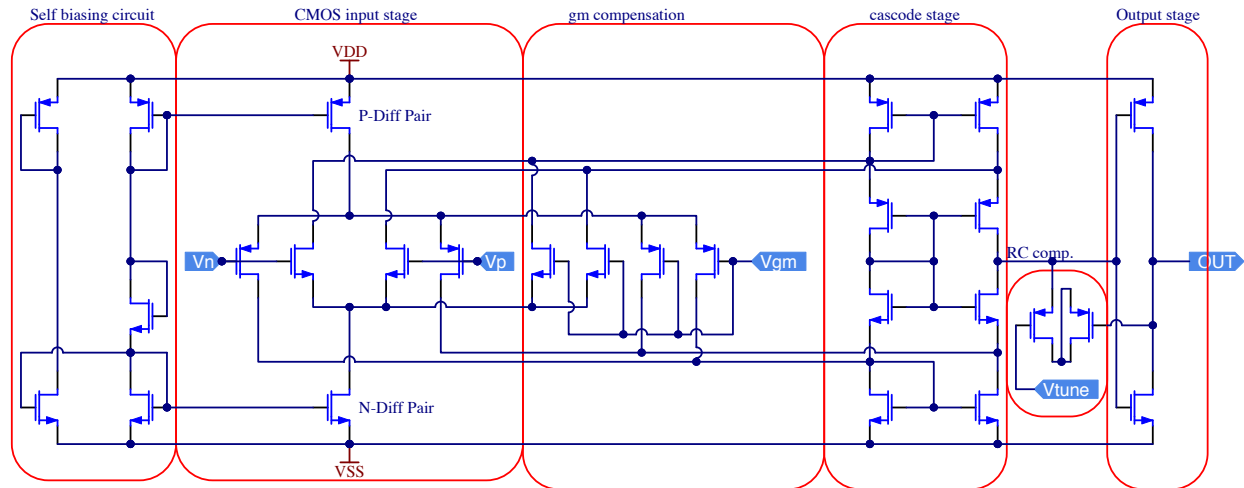


Figure 5.47: Schematic of the general purpose input/output rail-to-rail operational amplifier implementation.

inverter, which is configured as an AB class amplifier. The PMOS on the far right serves as the Miller compensation capacitor between the cascode and output stages. The op-amp is surrounded by a rectangular N-type guard ring, thus bringing the dimensions to $14.765 \mu m \times 8.83 \mu m$.

Figure 5.49a shows the input/output characteristic as a function of the input DC voltage in a unity gain buffer configuration. The deviation of the output versus the input is less than one mV between $0.1 V$ and $1.2 V$ and it rises to a maximum of $4 mV$ at the positive rail. Due to a very compact layout, the effects of mismatch are negligible. Figure 5.49b shows the transconductance of the first stage as a function of input DC voltage. The transconductance of the first stage defines the dominant pole of the input/output transfer function, thus a linear transconductance over the dynamic range simplifies the stabilization circuit. The g_m obtained is not perfectly linear over the entire dynamic range; however, it does provide with a significant improvement over a regular non-linearized response.

In terms of AC response, the open loop gain is $48.9dB$ or $278.3 V/V$. Figures 5.50a and 5.50a show the open loop response of the op-amp along with the phase margin as functions of the capacitive load and temperature. The op-amp stability is exceptional in this case. However, the unity gain closed-loop case sees a significant degradation in stability due to the output load capacitance.

Figure 5.51a shows the large signal bandwidth of the amplifier with a signal amplitude of $1 V_{PP}$. The unity gain buffer can thus drive a load capacitance of up to $3 pF$ at full swing for frequencies up to approximately $15 MHz$. The reader may notice that small capacitive loads do not result in higher bandwidths. This is because the frequency compensation circuit is set to maximize stability and it therefore slows down the op-amp even at very low capacitive loads.

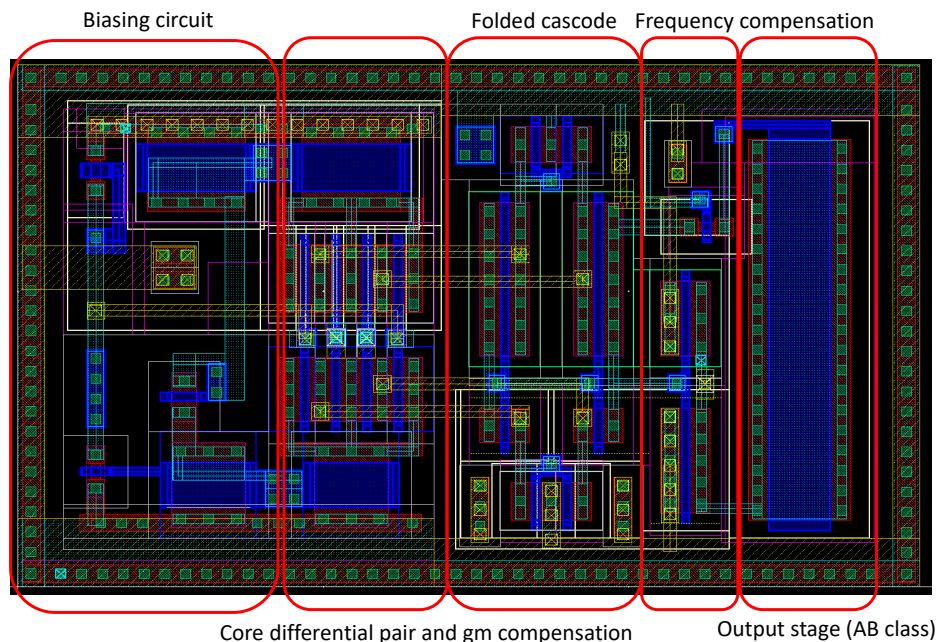


Figure 5.48: Layout of the general purpose input/output rail-to-rail operational amplifier implementation.

In the op-amp version with the tuning option, the bandwidth for low capacitive loads can be increased to up to 100 MHz .

Figures 5.52a and 5.52b show the input referred noise spectrum and the integrated noise floor as functions of the loading capacitance and temperature. As expected, the noise increases with temperature. The relationship with the load capacitance is a bit more complex. Up to approximately 5 pF , the integrated noise floor is almost flat, which is due to the fact that the bandwidth is limited and almost constant. Beyond 5 pF , the bandwidth increases because the amplifier becomes unstable and a resonance peak begins to appear, thus moving the 3 dB to higher frequencies. In addition, the resonance peak provides some noise gain. Obviously this region of operation is undesirable. In the region of operation of interest, the integrated noise floor is approximately $200\text{ }\mu\text{V}$, which is below one count of a 12-bit DAC.

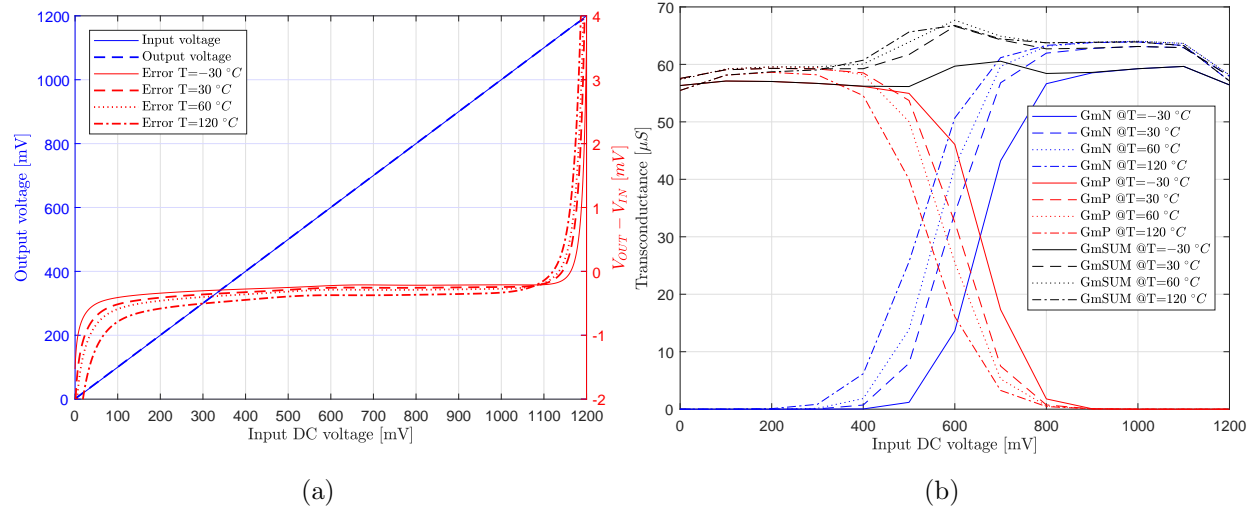


Figure 5.49: (a) Input/output characteristic as a function of the input DC voltage in a unity gain buffer configuration. (b) Transconductance as a function of input DC voltage.

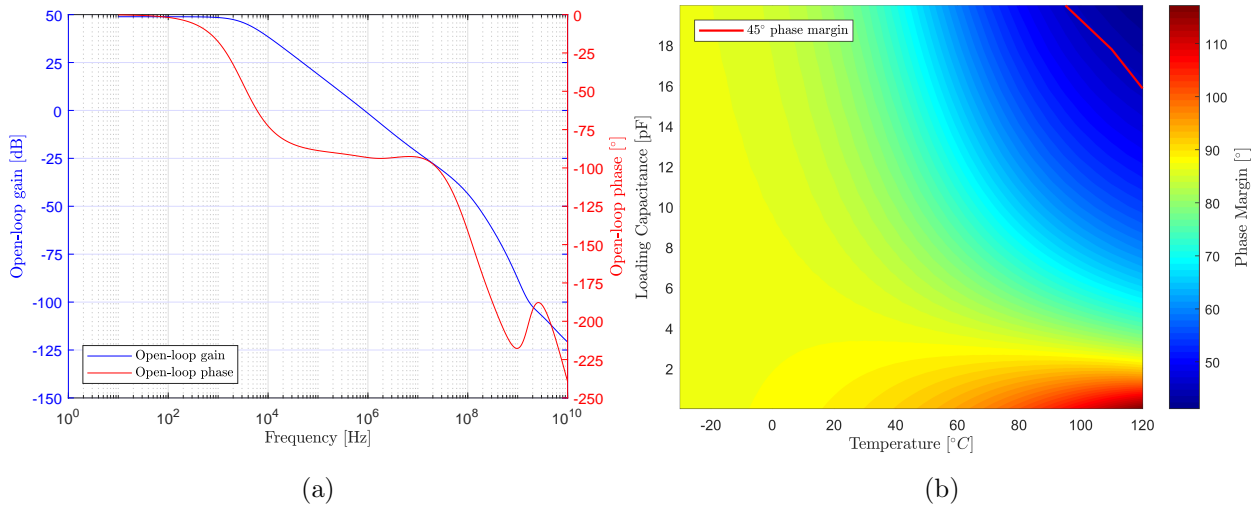


Figure 5.50: (a) Open-loop gain and phase response of the op-amp. (b) Open-loop phase margin as a function of load capacitance and temperature.

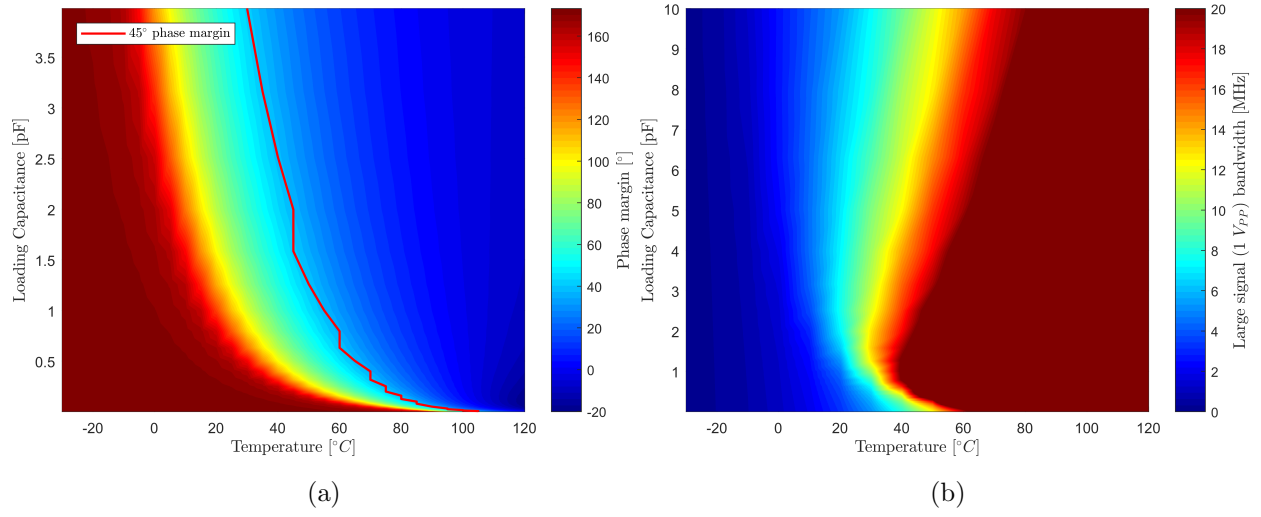


Figure 5.51: (a) Closed-loop phase margin of the op-amp in a unity gain configuration as a function of load capacitance and temperature. (b) Closed-loop large signal bandwidth as a function of load capacitance and temperature for the unity gain op-amp configuration.

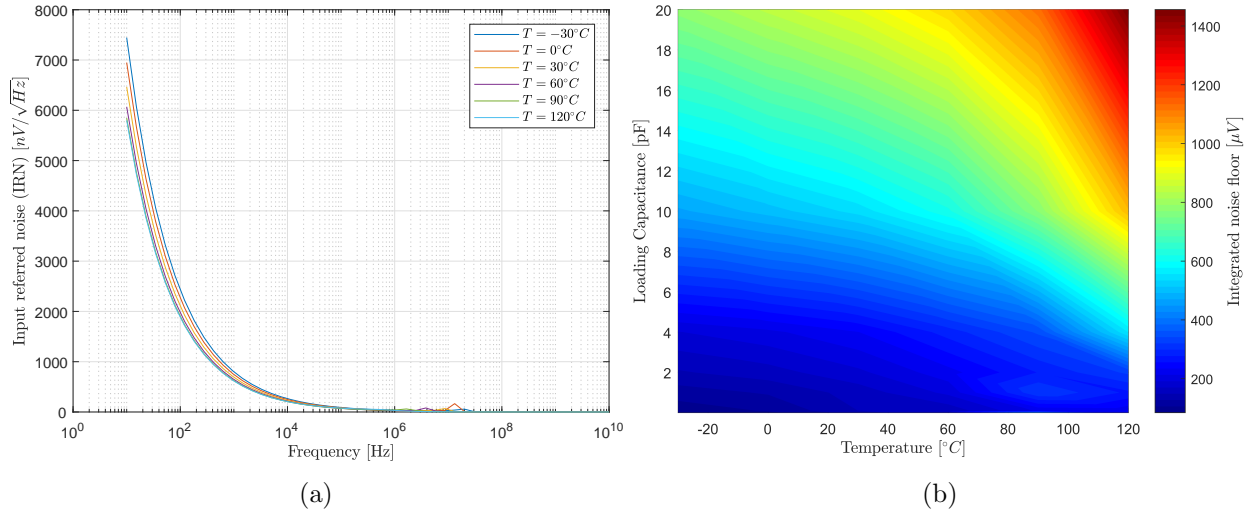


Figure 5.52: (a) Amplifier input referred noise for six different temperatures. (b) Amplifier integrated noise floor as a function of temperature and load capacitance.

5.6 Summary of the RFpix1 Analog Front-End

Figure 5.53 shows the RFpix1 analog front-end layout for two channels. The dimensions of the two channels are $2687.03 \mu m \times 340 \mu m$. The vertical dimension as seen in the figure has been chosen specifically to fit a pin pitch of $90 \mu m$, where each channel is connected to two pins. It can be observed that the input buffer is not included in this layout, because it has not been designed yet.

The total simulated current draw of a single channel is approximately $41.71 mA$, without considering the input buffer. The integrated noise floor of the analog front-end is $1.05 mV_{RMS}$, which translates to an ENOB of $9.6 bits$. This represents the achievable vertical resolution assuming an almost perfect calibration procedure. Figure 5.54 shows the simulated total error of the RFpix1 analog front-end.

The total sampling error is dominated by the non-linear response of the sampling buffer.

Table 5.2 summarizes and compares the simulated performance parameters of the RFpix1 analog front-end with the baseline RFpix1 design specifications.

Table 5.2: Simulated RFpix1 performance in comparison with the baseline specifications.

<i>Parameter</i>	<i>Desired value</i>	<i>Simulated value</i>
Sampling period	$50 ps @ 20 GS/s$	$50 ps @ 20 GS/s$
Analog bandwidth ^a	$\approx 3 GHz$	$\approx 3.56 GHz$
Input referred noise ^b	$\leq 0.5 mV_{RMS}$	$\approx 1.05 mV_{RMS}$
Added jitter per channel	$\approx 40 fs$	$\approx 29 fs$
ENOB ^c	≥ 10	≈ 9.6
Power consumption per channel ^b	$40 mA$	$41.71 mA$

^a The simulated value is the tracking bandwidth of the SCA.

^b The simulated value does not take into account the input buffer.

^c The simulated value does not take into account distortion.

The current prototype design is able to achieve very low added jitter, however the amplitude noise is twice as much as needed. By inserting this values into the synthetic waveform generator presented in chapter four we find that the achievable transition time resolution is approximately $160 fs$.

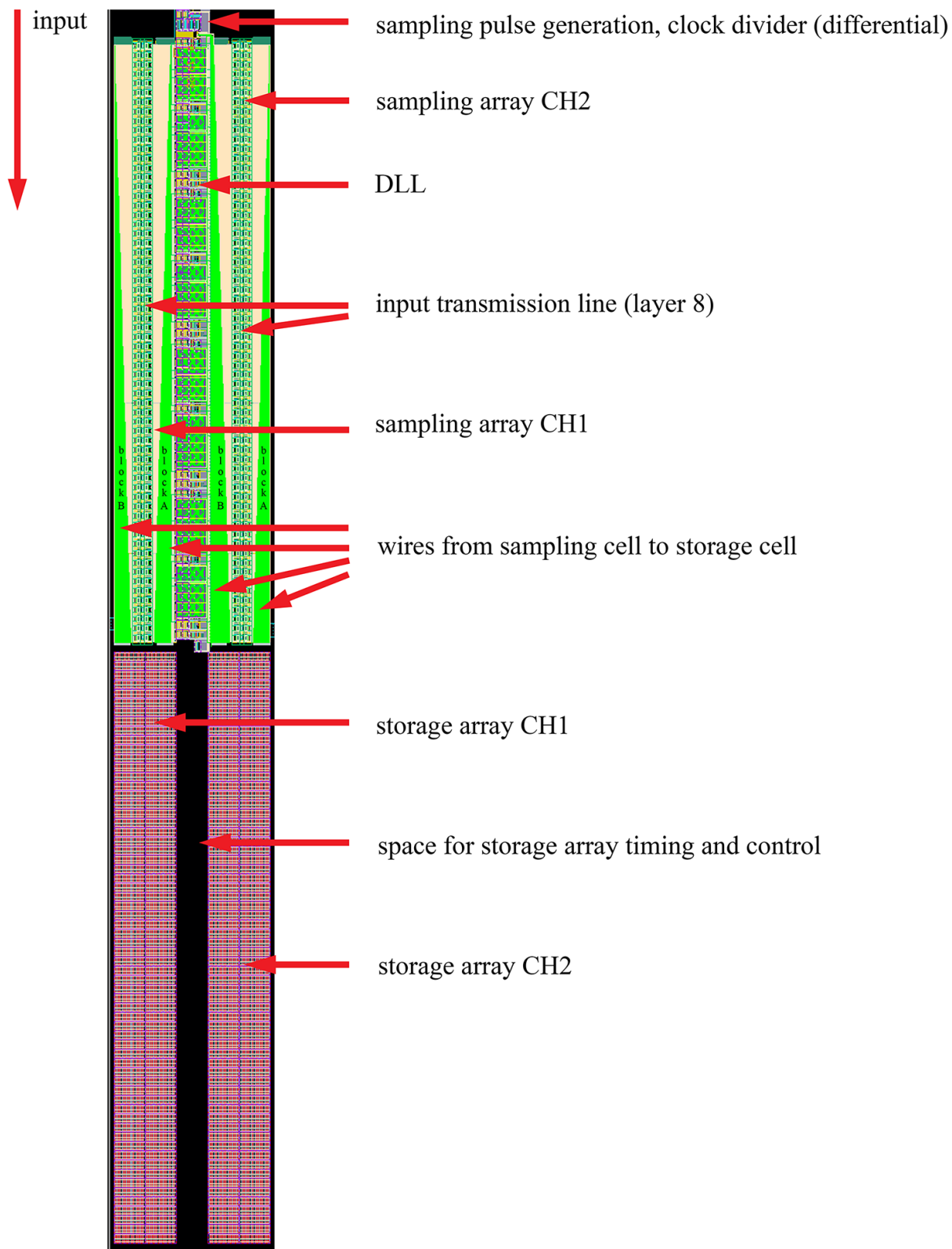


Figure 5.53: RFpix1 analog front-end layout.

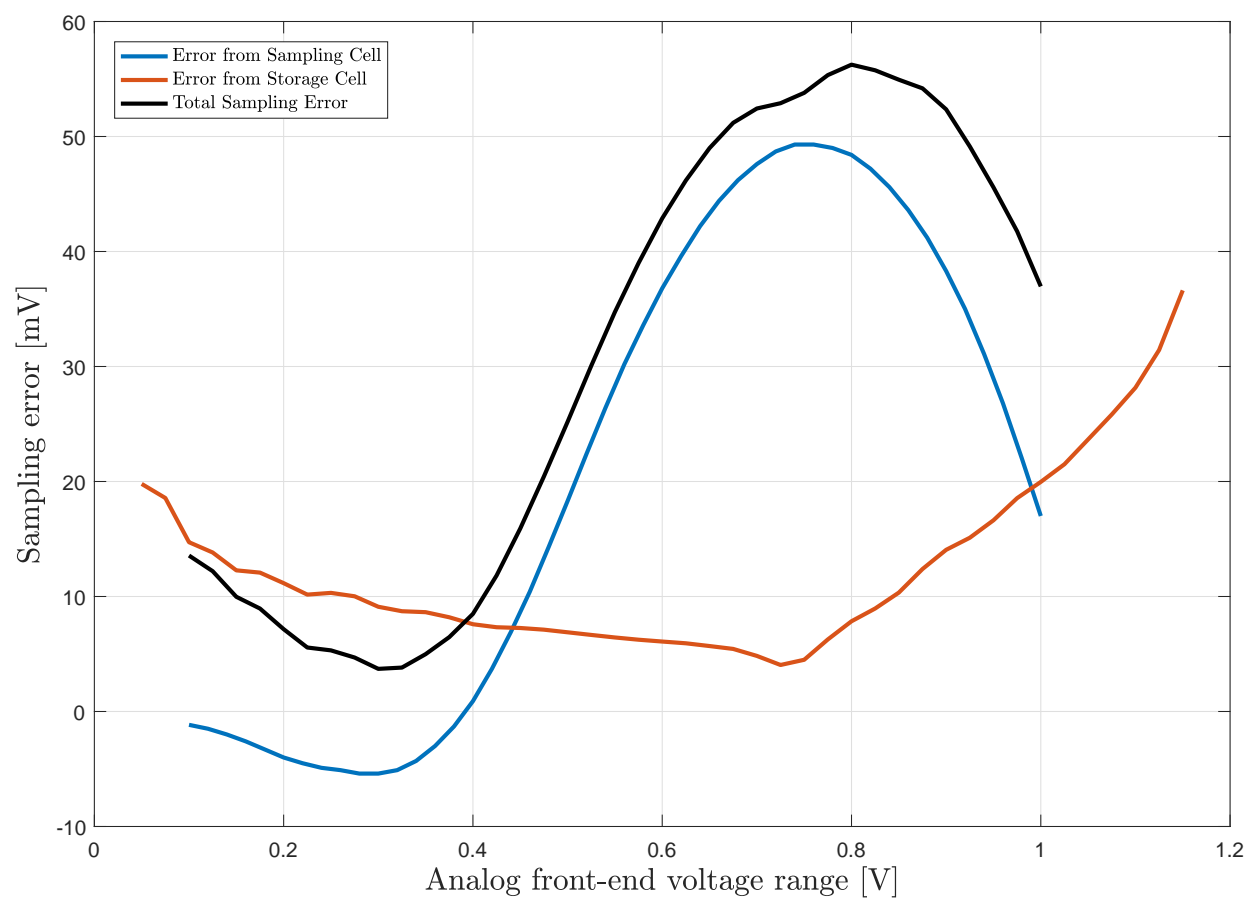


Figure 5.54: Total estimated sampling error of the RFpix1 analog front-end.

CHAPTER 6

SUMMARY

With a saturated maximum hit rate tolerance of $1.25 \cdot 10^{10} \text{ Hits/cm}^2\text{s}$, exquisite space-time resolution, relatively low power, and low data rate per event, the TVD concept presented is a viable upgrade to the current pixel detector of the Belle II spectrometer. In addition, the RFpix ASIC could be used for precision readouts of other strip line detectors, such as the Large Area Picosecond PhotoDetectors (LAPPD) [76], with femtosecond timing leading to enhanced spatial resolution. Furthermore, a similar sensor layout based on such a principle of operation could present a viable solution for vertex detectors in high luminosity environments of other present and future colliders like the LHC and the ILC. The TVD concept and its feasibility study results have been published in [75].

A synthetic waveform generator has been developed, validated through measurement results, and presented in chapter three. This software tool was crucial in studying, identifying, and quantifying the sources of error and their corresponding coupling mechanism that affect the extracted timing resolution at the femtosecond level for the waveform sampling technique. This study results were instrumental in setting the RFpix baseline specifications. The synthetic waveform generator and all of the study results have been published in [77].

The RFpix prototype (RFpix1) analog front-end has been designed in a 130 nm CMOS technology node. It is divided into three main sub-circuits: switched capacitor array with 128 sampling cells, a two-level delay-locked loop with very low added jitter (29 fs), and an analog storage array with the an integrated comparator, which is also a part of the single-slope ADC. All of the presented simulation results with layout parasitics, indicate that the RFpix1 can achieve the timing resolution of 160 fs with proper calibration. Such a timing resolution translates into a spatial resolution of 20.1 μm , which is good enough for a pixel grid with a 50 μm pitch. Most of this work has been published and/or presented in [78–80].

In order to complete the RFpix1 prototype, additional design work is required. More specifically:

- Analog front-end: some additional work is needed in terms of completing the input buffer design and the storage array control circuits;
- The single-slope ADC circuit with the ramp generator, counter and the register latches;
- The data transfer section with the serializer circuit and the LVDS drivers;
- The main clock distribution circuit;
- The digital control interface with the SPI, control registers, and DACs.

BIBLIOGRAPHY

- [1] J. L. Hewett and et al. Fundamental physics at the intensity frontier. In *2011 workshop on Fundamental Physics at the Intensity Frontier*, 2012.
- [2] Wikipedia. Standard model — wikipedia, the free encyclopedia, 2017. [Online; accessed 8-December-2017].
- [3] T. Abe and et al. Belle II Technical Design Report. *ArXiv e-prints*, November 2010.
- [4] C. Patrignani et al. (Particle Data Group). The review of particle physics (2017). *Chin. Phys. C*, 2017.
- [5] B. Kolbinger. Simulation of a silicon-strip detector. Technical report, HEPHY Institute of High Energy Physics, 2012.
- [6] Wikipedia. Babar experiment — wikipedia, the free encyclopedia, 2017. [Online; accessed 29-March-2017].
- [7] and J. Huijsing F. Witte, K. Makinwa. *Dynamic Offset Compensated CMOS Amplifiers*. Analog Circuits and Signal Processing. Springer Netherlands, 2009.
- [8] G. C. Branco and et al. Leptonic CP Violation. *Rev. Mod. Phys.*, 84:515–565, 2012.
- [9] Wikipedia. Superkekb — wikipedia, the free encyclopedia, 2017. [Online; accessed 22-November-2017].
- [10] Carl R. Nave. Electron volts, 2017.
- [11] H. Goldstein, C.P. Poole, and J.L. Safko. *Classical Mechanics*. Addison Wesley, 2002.
- [12] W. Herr and et al. *6.4 Concept of Luminosity*, pages 140–146. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [13] KEK HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION. SuperKEKB, 2011.
- [14] T. Kuhr and I. for the Belle Collaboration. Status of SuperKEKB and Belle II. *ArXiv e-prints*, January 2011.
- [15] Gary Jhon Barker. Silicon vertex detectors and particle identification. In *b-Quark Physics with the LEP Collider*, pages 23–35. Springer, 2010.
- [16] Norbert Wermes. Pixel Vertex Detectors. *ArXiv Physics e-prints*, November 2006.

- [17] E. Gatti and et al. Semiconductor drift chamber – an application of a novel charge transport scheme. *Nucl. Instrum. Meth. A*, 225(3):608 – 614, 1984.
- [18] Hiro Nakayama and for the Belle Collaboration. Beam background update 14th campaign, 2016. B2GM plenary session.
- [19] Carlos Marinas. The Belle II DEPFET vertex detector: Current status and future plans. *JINST*, 7(02):C02029, 2012.
- [20] Hatextyenkansson Per. An introduction to the time-of-flight technique. *Brazilian Journal of Physics*, 29:422 – 427, 09 1999.
- [21] Chris Robertson. *Printed Circuit Board Designer’s Reference*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2003.
- [22] W. Kester. Taking the mystery out of the infamous formula, “ $\text{SNR} = 6.02N + 1.76\text{dB}$,” and why you should care. Technical report, Analog Devices, 2009.
- [23] J.F. Genat and et al. Signal processing for picosecond resolution timing measurements. *Nucl. Instrum. Meth. A*, 607(2):387 – 393, 2009.
- [24] D. Stricker-Shaver and et al. Novel calibration method for switched capacitor arrays enables time measurements with sub-picosecond resolution. *IEEE Trans. Nucl. Sci.*, 61(6):3607–3617, Dec 2014.
- [25] M.L. Minges and A.S.M.I.H. Committee. *Electronic Materials Handbook: Packaging*. Electronic Materials Handbook. Taylor & Francis, 1989.
- [26] Pinakpani Nayak. *Characterization of High-Resistivity Silicon Bulk and Silicon-on-Insulator Wafers*. PhD thesis, Arizona state university, 2012.
- [27] Leif Jensen. High resisitvity (hirestm) silicon for GHz & THz technology. Technical report, TTopsil Semiconductor Materials A/S, 2014.
- [28] Rogers Corporation. *TMM Thermoset Microwave Materials*, 2015.
- [29] Dan McMahonill. Microstrip analysis/synthesis calculator, 2009.
- [30] Keysight Technologies. Advanced design system (ads), 2016. [Online; accessed 21-December-2016].
- [31] E. Rubiola. *Phase Noise and Frequency Stability in Oscillators*. Cambridge University Press, Published in the United States of America by Cambridge University Press, New York, 2008.

- [32] Pavel Buzhan and et al. The Advanced Study of Silicon Photomultiplier. In *Advanced Technology - Particle Physics*, pages 717–728, November 2002.
- [33] Giulio Pellegrini and et al. Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications. *Nucl. Instrum. Meth. A*, 765:12 – 16, 2014.
- [34] Joern Lange and et al. 3d silicon pixel detectors for the high-luminosity lhc. *JINST*, 11(11):C11024, 2016.
- [35] Gian Franco Dalla Betta and et al. Development of a new generation of 3d pixel sensors for hl-lhc. *Nucl. Instrum. Meth. A*, 824:386 – 387, 2016. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.
- [36] Valeri Saveliev. Avalanche pixel sensors and related methods, September 18 2012. US Patent 8,269,181.
- [37] I Tapan and et al. Avalanche photodiodes as proportional particle detectors. *Nucl. Instrum. Meth. A*, 388(1):79 – 90, 1997.
- [38] N. Cartiglia and et al. Beam test results of a 16 ps timing system based on ultra-fast silicon detectors. *Nucl. Instrum. Meth. A*, 850:83 – 88, 2017.
- [39] Giulio Pellegrini and et al. Recent technological developments on {LGAD} and ilgad detectors for tracking and timing applications. *Nucl. Instrum. Meth. A*, 831:24 – 28, 2016.
- [40] Nicola D’Ascenzo and et al. Silicon avalanche pixel sensor for high precision tracking. *JINST*, 9(03):C03027, 2014.
- [41] Eva Vilella Figueras. *Feasibility of Geiger-mode avalanche photodiodes in CMOS standard technologies for tracker detectors*. PhD thesis, Universitat de Barcelona, 2013.
- [42] Madhavan Swaminathan and Ki Jin Han. *Design and Modeling for 3D ICs and Interposers*. WSPC Series in Advanced Integration and Packaging. World Scientific Publishing Company, 2013.
- [43] E. Oberla and et al. A 15 GSa/s, 1.5 GHz bandwidth waveform digitizing ASIC. *Nucl. Instrum. Meth. A*, 735:452 – 461, 2014.
- [44] Mike Cooney and et al. Multipurpose test structures and process characterization using 0.13 μm CMOS: The CHAMP ASIC. *Physics Procedia*, 37:1699 – 1706, 2012. Proceedings of TIPP 2011.

- [45] Gary S. Varner and et al. The large analog bandwidth recorder and digitizer with ordered readout (labrador) {ASIC}. *Nucl. Instrum. Meth. A*, 583(2–3):447 – 460, 2007.
- [46] Synergy Microwave Corp. Ultra low-noise crystal oscillators. *Microw. J.*, April 2012.
- [47] Angelo Rivetti. *CMOS: Front-End Electronics for Radiation Sensors*. Devices, Circuits, and Systems. Taylor & Francis, 2015.
- [48] M.J.M. Pelgrom. *Analog-to-Digital Conversion*. Springer Netherlands, 2010.
- [49] B. Razavi. *Design of Analog CMOS Integrated Circuits, Second Edition*. McGraw-Hill Higher Education, 2016.
- [50] Wikipedia. Johnsonnyquist noise — wikipedia, the free encyclopedia, 2017. [Online; accessed 5-April-2017].
- [51] Walt Kester. Understand SINAD, ENOB, SNR, THD, THD + N, and SFDR so you don’t get lost in the noise floor. Technical report, Analog Devices, 2009.
- [52] B. Razavi. *Principles of data conversion system design*. IEEE Press, 1995.
- [53] P. Gill. Ultrafast optics: Femtosecond timing distribution. *Nature Photonics*, 2:711 – 712, 2007.
- [54] P. L. Lemut and et al. Chasing Femtoseconds, How Accelerators Can Benefit from Economies of Scale in Other Industries. *Conf. Proc.*, C110904:1973–1977, 2011.
- [55] P. Orel and et al. Next generation CW reference clock transfer system with femtosecond stability. In *Proceedings of PAC2013, Pasadena, CA USA*, pages 1358–1360, October 2013.
- [56] M. Collins and et al. On-chip timing measurement architecture with femtosecond resolution. *Electronics Letters*, 42(9):528–530, April 2006.
- [57] E. Bogatin. *Signal and Power Integrity: Simplified - 2nd edition*. Prentice Hall, Inc, 2010.
- [58] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.
- [59] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 2016.
- [60] F. N. Fritsch and et al. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.
- [61] MathWorks. Cubic spline data interpolation, 2017.
- [62] MathWorks. Piecewise cubic hermite interpolating polynomial (pchip), 2017.

- [63] Agilent/Keysight. Agilent e4428c esg analog signal generator data sheet, 2012.
- [64] Texas Instruments. Adc12j4000evm user’s guide, 2015.
- [65] LeCroy. LeCroy sda 13000 data sheet, 2007.
- [66] Tektronix. Digital storage oscilloscope tds6000b/c series data sheet.
- [67] Agilent/Keysight. Agilent infiniiium 8000 series oscilloscopes data sheet, 2013.
- [68] Texas Instruments. ADC12J4000 12-Bit 4 GSPS adc with integrated DDC, 2015.
- [69] B. S. Everit and et al. *The Cambridge Dictionary OF Statistics - 4nd edition*. Cambridge University Press, 2010.
- [70] Wikipedia. Mexican hat wavelet — wikipedia, the free encyclopedia, 2017.
- [71] M. Figueiredo and R. L. Aguiar. Predicting noise and jitter in cmos inverters. In *2007 Ph.D Research in Microelectronics and Electronics Conference*, pages 21–24, July 2007.
- [72] J. Huijsing. *Operational Amplifiers: Theory and Design*. Springer International Publishing, 2016.
- [73] M.L. Boas. *Mathematical Methods in the Physical Sciences*. Wiley, 2005.
- [74] Tina Harriet Smilkstein. *Jitter Reduction on High-Speed Clock Signals*. PhD thesis, University of California at Berkeley, 2007.
- [75] Peter Orel, Gary S. Varner, and Pardis Niknejadi. Exploratory study of a novel low occupancy vertex detector architecture based on high precision timing for high luminosity particle colliders. *Nucl. Instrum. Meth.*, A857:31–41, 2017.
- [76] Bernhard Adams and et al. Measurements of the gain, time resolution, and spatial resolution of a $20 \times 20 \text{ cm}^2$ mcp-based picosecond photo-detector. *Nucl. Instrum. Meth. A*, A732:392–396, 2013.
- [77] P. Orel and G. S. Varner. Femtosecond resolution timing in multi-gs/s waveform digitizing asics. *IEEE Transactions on Nuclear Science*, 64(7):1950–1962, July 2017.
- [78] P. Orel and G. S. Varner. Improved switched capacitor cell for a 3 ghz 20 gs/s waveform digitizing asic. In *2017 13th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, pages 129–132, June 2017.
- [79] P. Orel and G. S. Varner. Development of a waveform sampling asic with femtosecond timing for a low occupancy vertex detector. In *2017 Topical Workshop on Electronics for Particle Physics (TWEPP)*, September 2017.

- [80] P. Orel and G. S. Varner. Two-level dll with femtosecond added jitter for a low power 20 gs/s sampling asic. In *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC)*, October 2017.